А. П. ХАУСТОВ

Многомерный АНАЛИЗ гидрогеологических СИСТЕМ горно-складчатых ОБЛАСТЕЙ

$$\mathrm{Cos}\,\theta_{ij} = \frac{\sum_{k=1}^{m} x_{i,k} x_{j,k}}{\sqrt{\sum_{k=1}^{m} x_i^2,\ k \sum_{k=1}^{m} x_{j,k}^2}}$$

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ЗЕМНОЙ КОРЫ

А.П. ХАУСТОВ

МНОГОМЕРНЫЙ АНАЛИЗ ГИДРОГЕОЛОГИЧЕСКИХ СИСТЕМ ГОРНО-СКЛАДЧАТЫХ ОБЛАСТЕЙ

Ответственный редактор
д-р геол.-мин. наук И. С. Ломоносов

4706

НОВОСИБИРСК
ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ
1986

Институт геологических наук АН Армянской ССР
Библиотека

УДК 556.0.13 (571.53/55)

Хаустов А.П. Многомерный анализ гидрогеологических систем горно-складчатых областей. – Новосибирск: Наука, 1986.

В монографии анализируются методологические аспекты применения комплекса методов многомерного анализа (факторного, кластерного, многомерных полей) при изучении процессов формирования ресурсов подземных вод горно-складчатых областей (на примере бассейна оз. Байкал). С позиций системного подхода рассмотрены сложные взаимосвязи элементов гидрогеологических систем. В качестве основного аппарата познания процессов формирования ресурсов подземных вод используются многомерные классификации. Показаны логические процедуры классификаций, методы расчета и отбора информативных признаков, проведены оценка работы соответствующих алгоритмов и интерпретация конечных результатов классификации, а также выявлены возможности их практического применения при среднемасштабном гидрогеологическом картировании и прогнозе количественных показателей естественных ресурсов подземных вод по комплексу косвенных данных.

Книга является методически новой и представляет интерес для гидрогеологов, гидрологов, специалистов водного хозяйства и лиц, работающих в области применения математических методов в естественных науках.

Рецензенты С.М. Чесалов, А.А. Дзюба

X $\frac{1904060000-861}{042(02)-86}$ 185-86-IV

© Издательство "Наука", 1986 г.

ОТ РЕДАКТОРА

Современный прогресс естественных наук во многом определяется широким внедрением системного принципа изучения природных процессов и электронно-вычислительной техники, предопределяющих возможность комплексирования различных методов моделирования природных систем, объединенных общей идеей или общим принципом познания действительности. Системный принцип прочно вошел в физические, технические, биологические и общественные отрасли знаний. Создано новое научное направление - системология.

Основная функция системного анализа - целенаправленная переработка исходной информации с целью получения новых сведений об изучаемых процессах и явлениях. Его практическое применение возможно лишь в случае, когда определены границы объекта исследований, изучена структура связей между основными свойствами объекта и его элементами, т.е. существует концептуальная модель интересующего явления. Отсутствие формализации понятийной базы геологии долгое время являлось серьезным препятствием на пути ее математизации. Особенно эффективно применение принципов системного подхода при гидрогеологических исследованиях, так как современный уровень знаний о процессах, протекающих в подземной гидросфере, позволяет создать практически любую концептуальную модель, в той или иной степени адекватную реальным условиям, и, следовательно, имеются практически неограниченные возможности моделирования этих процессов в рамках системного подхода.

В монографии А.П. Хаустова в качестве основного аппарата исследования гидрогеологических систем используются таксономические модели условий формирования естественных ресурсов подземных вод бассейна оз. Байкал, основанные на широком применении методов статистического многомерного анализа (факторного, компонентного, кластерного, регрессионного и др.). Поскольку формирование ресурсов подземных вод происходит под влиянием целого комплекса природных условий, приходится иметь дело с многомерной структурой эмпирических данных. Следовательно, применение мощного статистического аппарата оправданно и необходимо при решении задач таксономии.

Озеро Байкал - весьма удачный объект исследования, поскольку многообразие процессов его геологического развития обусловило раскрытость гидрогеологических систем, а пространственная изменчивость параметров ресурсов подземных вод региона во многом отражает новейшие движения земной коры в условиях рифтогенных структур. Кроме того, выбранный объект отвечает одному из основных требований системного анализе - целостности системы. При построении таксономических моделей и анализа комплекса взаимосвязи гидрогеологических систем автором аргументированно используются такие показатели, как градиенты вертикальных тектонических движений земной коры и показатели плотности тектонических нарушений.

Сложность процессов формирования подземных вод требует от исследователя изучения большого числа гидрогеологических объектов и рассмотрения

достаточного набора признаков, вскрывающих основные закономерности процессов. Связь признаков между собой отражает взаимодействие процессов; именно поэтому в основу предлагаемой монографии положено изучение связей между переменными с помощью методов многомерного анализа. Практическая реализация алгоритмов многомерного анализа требует обращения к большим ЭВМ, что в известной мере сдерживает широкое его применение, однако получаемый эффект покрывает издержки затраченного времени на подготовку исходных данных.

Гидрогеологические классификации как метод обобщения информации создаются со времен становления науки, хотя методологические основы их построения разработаны слабо. При создании большинства классификаций специалист прежде всего опирается на свой опыт и интуицию, что придает классификациям определенную долю субъективизма и снижает их научную и практическую ценность. Следует добавить, что возможности человека в этой области весьма ограниченны и позволяют учесть незначительное число признаков и объектов, в то время как возможности больших ЭВМ практически неограниченны.

В последние годы в практике изучения природных систем широко применяются формальные методы построения типологии, с помощью которых исходная совокупность эмпирических данных по одному или нескольким признакам распределяется на однотипные группы или классы. Современные требования к качеству гидрогеологических классификаций, а следовательно, и к гидрогеологическому районированию предполагают учет всех значащих признаков. При этом весьма важен вопрос о числе объектов и признаков, используемых для построения классификаций: малое число признаков и объектов приводит к неустойчивости классификационных схем, большое количество требует применения автоматических методов классификации. Любые различия классификаций порождают многообразие теоретических построений, и следовательно, их неопределенность.

В монографии А.П. Хаустова предлагается новый методический подход к исследованию формирования ресурсов подземных вод горно-складчатых областей, при этом многомерные классификации выступают в качестве основного аппарата познания.

И.С. Ломоносов

Глава 1

СИСТЕМНЫЙ ПОДХОД И МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ КЛАССИФИКАЦИИ ГИДРОГЕОЛОГИЧЕСКИХ СИСТЕМ

1.1. СИСТЕМНО-КИБЕРНЕТИЧЕСКИЙ ПРИНЦИП ПРЕДСТАВЛЕНИЯ ГИДРОГЕОЛОГИЧЕСКИХ СИСТЕМ

В настоящее время методы системного анализа довольно широко применяются в естественных науках. Аспекты системного подхода при решении целого комплекса научных и практических задач, связанных с изучением подземных водных ресурсов, рассматривались в работах А.Г. Булавко /1971/, И.К. Гавич /1980, 1984/, В.Т. Дубинчука /1982/, К.П. Караванова /1977, 1980/, L. Luckner /1975/, Н.А. Огильви /1974/, А.Н. Павлова /1981/, Е.В. Пиннекера /1977/, Н.В. Роговской /1981/, У. Умарова /1982/, Н.М. Фролова /1981, 1984/, П.Ф. Швецова, П.А. Киселева/, 1979/, П.Ф. Швецова /1984/ и др.

На наш взгляд, наиболее удачное определение системного анализа дано в работе Дж. Джефферса /1981/, который под системным анализом понимает не какой-то один метод или группу, а прежде всего широкую стратегию поиска, базирующуюся на математическом аппарате и математической концепции, но в рамках систематизированного подхода к решению сложных проблем. Системный анализ - это прежде всего упорядоченное и логическое представление данных в виде моделей, но необходимо помнить, что всякая система есть модель, но не всякая модель есть система. Построение моделей - необходимый, но не исчерпывающий признак системного подхода. Авторами одной из работ дается следующее категоричное его определение: «...это не наука и не теория, а образ мышления, методологическая концепция, исследовательская точка зрения, позиция или системная ориентация...» /Методы..., 1978, с. 25/.

Использование системных принципов в гидрогеологических исследованиях вызвано необходимостью глубокого изучения внутренних и внешних свойств подземной гидросферы, выявления присущих только ей специфических закономерностей, которые не могут быть выведены из закономерностей относительно простых и изученных элементов. Сказанное достаточно емко сформулировано в одной из фундаментальных теорем К. Геделя: в любой формальной системе понятий возникают задачи, которые не могут быть решены иначе, как путем расширения самой системы понятий /Кондаков, 1971/.

В системных исследованиях достаточно конкретно сформировалось два направления: общая теория систем и частные научные разработки в рамках системно-структурного анализа /Садовский, 1974/.

Сравнительно недавно Б.Г. Юдиным /1972/ был предложен термин "системная ориентация". По-видимому, этот термин наиболее точно отражает современное состояние системного движения в гидрогеологии, поскольку сложившаяся ситуация на данном этапе испытывает настоятельную потребность широкого внедрения методов системного анализа.

Общая теория систем как межнаучная дисциплина к настоящему времени уже сформировалась, и более перспективным в гидрогеологических исследова-

ниях представляется обратный путь – поиск и определение системных свойств на конкретных объектах и процессах формирования подземных вод.

При широком внедрении методов системного анализа в гидрогеологии возникает ряд трудностей. Одна из основных трудностей – широкий охват гидрогеологических проблем и, как следствие этого, несовершенство терминологической базы и понятий в терминологии системной концепции. Зачастую одни и те же проблемы с позиций системного анализа трактуются противоречиво, причем эти позиции, как правило, заимствованы из других наук (техники, биологии, философии и др.). Хотя системный анализ и призван упорядочить представления о ряде гидрогеологических проблем, до сих пор мы не можем выйти за рамки традиционных методов решения этих проблем. Нельзя не упомянуть еще об одной трудности на пути внедрения идей системного анализа. Поскольку последние служат отправным пунктом методологического обоснования математического моделирования гидрогеологических систем, а модели в значительной мере строятся на формализации их свойств, создается тенденция пренебрежения сложностью реальных гидрогеологических процессов, многие из которых настолько сложны, что их упрощение и схематизация отделяют построенные модели от реального мира, делают непригодными для практических исследований или искажают действительность.

Гармоничную стыковку системного подхода с традиционными методами гидрогеологических исследований можно осуществить с помощью принципов обратной связи, рассматривая на каждой стадии познания гидрогеологических систем парадигмы системного подхода через реальность гидрогеологических процессов.

В начале системных исследований в геологических науках было выделено три класса геологических систем: статические, динамические и ретроспективные /Косыгин, Соловьев, 1969/. В работах же по общей теории систем выделяются и функциональные, поэтому, рассматривая подземные водоносные системы как динамические, хотя своим существованием они обязаны статическим и ретроспективным системам, наиболее верным в методологическом отношении является комплексное рассмотрение всех трех или даже четырех разновидностей систем. В качестве концепта, связывающего разновидности систем и их элементы, можно рассматривать (с точки зрения поставленных задач) подземные воды.

Существует довольно обширный список характеристик и свойств разновидностей систем /Уемов, 1978; Забродин, 1981б/. Не претендуя на полноту охвата этого списка, остановимся лишь на некоторых, наиболее важных с позиций обоснования математического аппарата исследований, используемого в предлагаемой монографии.

Целостность систем обеспечивается тем, что сила связи между ее элементами выше, чем между системой и средой /Забродин, 1981б/. Это свойство системы выступает наряду с такими самостоятельными свойствами, как сложность, разнообразие, структура, иерархичность, гомогенность и гетерогенность, аддитивность и эмерджентность, открытость и закрытость, дискретность и непрерывность и др. Целостность системы предусматривает необходимость изучения свойств ее составных частей во взаимной связи и предполагает структурную сложность, гетерогенность и возможность моделирования отдельных частей системы и их последующее рассмотрение в качестве относительно независимых образований. Наглядное представление этапов системного подхода к изучению формирования ресурсов подземных вод удобно проиллюстрировать схемой (рис. 1). Выделенные этапы – не предел и не догма, их может быть (на определенной стадии исследований) больше или, наоборот, мень-

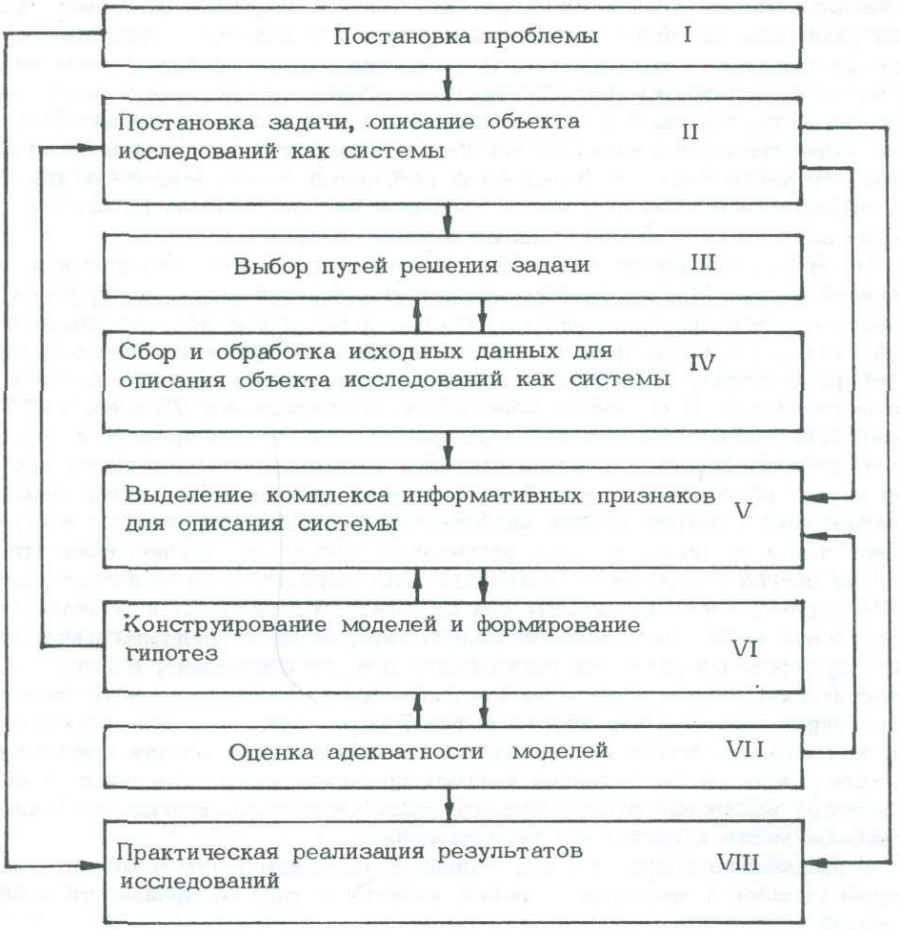


Рис. 1. Блок-схема основных этапов системного анализа изучения процессов формирования ресурсов подземных вод.

ше. Это же относится к взаимодействию между блоками на разных этапах системного подхода.

На первом этапе требуется определить цель исследований, необходимость применения системного анализа, его эффективность, реальные возможности и технические средства коллектива или индивидуума. Эту проблему в гидрогеологии рассматривают как выработку концепций, например связанных с оценкой ресурсов подземных вод для достаточно крупных регионов. Необходимо также определить объект исследований и его пространственно-временные границы. Например, для Восточной Сибири это может быть Сибирская платформа, или конкретные артезианские бассейны, или Байкальская рифтовая зона. Конкретизацией объектов исследования обеспечивается интерпретация моделирования гидрогеологических процессов.

Исходя из задач исследования любая проблема может быть условно разделена на более мелкие (например, в региональном плане) или на две более простые задачи, каждая из которых преследует свою цель. Для платформен-

ных областей оценка естественных ресурсов подземных вод может быть проведена раздельно как для зоны интенсивного, так и для зон затрудненного и весьма затрудненного водообменов. В горно-складчатых областях такие же задачи могут решаться отдельно для гидрогеологических массивов и межгорных артезианских бассейнов. Иными словами, на втором этапе исследований необходимо выделить исследуемый объект из более широкой системы, включающей его как составную часть. В дальнейшем выбранный объект выступает как область ограниченного и непрерывного пространства, выделяемая на основе единства внутренних и внешних связей элементов системы.

На третьем этапе исследований необходимо наметить пути решения поставленной задачи, что во многом зависит от специфичности моделируемых процессов и объектов. Чаще всего проблемы, аналогичные поставленным в данной работе, решались традиционными методами (в предположении о некотором наборе факторов, влияющих на основные закономерности распределения подземного стока). В последние годы рядом исследователей /Зекцер, 1977; Ковалевский, 1976, 1983; и др./ предприняты попытки отобразить количественно меру воздействия отдельных факторов на пространственно-временную изменчивость подземного стока. Для этих целей, как правило, использовались корреляционный и регрессионный анализы, общий недостаток которых - ограниченное число вводимых в схему расчетов независимых предикторов, что далеко не всегда обеспечивает наглядную достоверность регрессионных моделей. Расширение числа возможных связей между переменными в моделях при их построении может дать дополнительную информацию о предполагаемых процессах формирования ресурсов подземных вод и, следовательно, повысить их прогностическую основу. Здесь мы сталкиваемся с одним из свойств системее структурностью. Под структурными понимаются связи, характеризующие систему как взаимодействие ее элементов. На основе установления функциональных связей может быть выявлена взаимокоррелированность отдельных элементов, поэтому выявление функциональных связей в системе всегда занимало центральное место в системных исследованиях.

Формирование подземных вод - процесс случайный, находящийся под влиянием комплекса природных условий, которые трудно воспроизвести в лабораторной обстановке.

Исходя из многообразия выборки и признакового пространства можно прийти к выводу о существовании устойчивых гидрогеологических процессов и связей, которые в дальнейшем можно идентифицировать. Предполагается, что наблюдения и результаты их анализа (см. гл. 2) рассматриваются относительно более широкой системы признаков, которая не поддается измерению полностью. В общем виде система признаков объектов может быть разделена на признаки, характеризующие как более общие черты, так и индивидуальные.

При изучении формирования подземных вод в горно-складчатых областях зачастую трудно четко выделить отдельные процессы, поскольку мы имеем дело с высокой степенью их динамичности. Такие системы (по В.В. Налимову /1971/) носят название диффузных, плохо организованных, в которых нельзя выделить и наблюдать в чистом виде процессы одной физической природы, зависящие от малого числа признаков. Для описания подобных систем необходимо большое число переменных (признаков). Можно предположить, что бóльшую информацию о поведении диффузных систем несут признаки, испытывающие интенсивную вариацию во времени и пространстве; именно на них необходимо обращать усиленное внимание при моделировании гидрогеологических процессов.

М.Д. Белонин с соавторами /1982/ справедливо указывали, что при изучении диффузных моделей наиболее эффективны методы многомерной стати-

стики. Поэтому вероятностно-статистическое моделирование – обязательный этап познания данной разновидности систем. Рекомендуется на начальной стадии моделирования подвергать анализу как можно большее число признаков с целью адекватного описания системы.

В настоящее время существует свыше 40 словесных и формализованных понятий-определений "система". Наиболее сжатое определение можно найти у одного из основателей системного анализа Л. Берталанфи /Садовский, 1974/. Под системой он понимает комплекс элементов и их связей, находящийся во взаимодействии и выступающий в заданном отношении как целое. Существует некоторый инвариант термина "система".

1. Система – целостный комплекс взаимосвязанных элементов.

2. Элемент системы не может быть описан без учета его места в системе.

3. Система образует особое единство со средой.

4. Свойства системы порождаются из свойств элементов, и наоборот, свойства элементов порождаются из характеристики целого.

5. Любая система является подсистемой, состоящей из элементов системы более высокого порядка. Подсистема, в свою очередь, является системой, но более низких порядков /Садовский, 1974; Уемов, 1978/.

В отечественной гидрогеологии определение гидрогеологических систем также давалось неоднократно.

И.К. Гавич /1984/ под гидрогеологической системой понимает упорядоченный комплекс гидрогеологических элементов, объединенных в единое целое, с определенными свойствами и отношениями между собой и внешней средой. Основными характеристиками любой гидрогеологической системы, в понимании И.К. Гавич, являются: свойства, структура, характер связей и взаимодействия между элементами системы и внешней средой, внутренние и внешние границы, которые разделяют систему на элементы и отделяют от внешней среды.

Н.М. Фролов /1981/ считает, что гидрогеотермодинамическая система есть водоносный комплекс или серия комплексов, характеризующихся более прочными внутренними связями, чем внешними, общностью истории гидрогеологического развития, проявляющейся в индивидуальности и обособленности тепло-, массообмена. Позже Н.М. Фролов /1984/ рассматривает гидрогеотермодинамические системы как элементы системы гидрогеосферы: под последними понимается совокупность взаимосвязанных элементов, главным из которых являются подземные воды.

П.Ф. Швецов /1984/ под тепло- и водообменной литосистемой понимает упорядоченное в той или иной мере целостное множество взаимодействующих компонентов с набором сильных связей между ними и их свойствами. К основным компонентам отнесены: твердая (минеральная) фаза, водные растворы и газы.

Н.А. Огильви /1974/ и К.П. Караванов /1980/ используют термин "подземная водоносная система" как синоним гидрогеологической или водоносной системы, понимая под этим участок земной коры, состоящий из гидрогеологических элементов, выделенных по какому-либо свойству и обладающих целостной структурой.

Таким образом, приведенные определения гидрогеологической системы логически не противоречат понятиям общей теории систем.

Последующие этапы системного подхода базируются на представлении о речном водосборе или его части как о многопараметрической динамической системе, в которой могут быть выделены взаимосвязанные элементы различных структурных уровней. Такой подход позволяет использовать кибернетиче-

ские принципы исследования закономерностей формирования подземных и поверхностных водных ресурсов. Основной функцией системы, с позиции изучения формирования водных ресурсов, является трансформация атмосферных осадков в другие составляющие водного баланса.

Речной водосбор имеет четко выраженные границы (водораздел), за исключением случаев несовпадения поверхностных и подземных площадей; разветвленную речную сеть, через которую осуществляется взаимосвязь поверхностных и подземных вод; ряд других элементов, так или иначе связанных между собой и являющихся отличительными чертами существования системы. Подземный и поверхностный сток в замыкающем створе - интегральная характеристика результата взаимодействия комплекса природных элементов.

С кибернетических позиций любое исследование связано с установлением зависимости воздействие - отклик. Воздействие подается на "вход", отклик фиксируется на "выходе" (схема черного ящика Н. Винера). Если принять за "вход" количество атмосферных осадков, а на "выходе" фиксировать параметры подземного стока, то в качестве внутренних управляющих элементов можно принять такие признаки, как степень проницаемости горных пород, испарение с водосбора и поверхности грунтовых вод, уклон местности, показатели вертикального и горизонтального расчленения рельефа и другие признаки. Общая зависимость между состоянием "входа" (X) и "выхода" (У) может быть задана при помощи переходной функции

$$X = R(Y),$$

где R - оператор преобразования - может быть представлен в виде и детерминированных, и стохастических моделей и является функциональным описанием системы. Это преобразование может быть задано логическим, алгебраическим, дифференциальным операторами; скалярными, векторными и матричными формами, составленными на основании знания устройства системы /Дружинин, Конторов, 1976/.

Для природных систем механизм R-преобразования, как правило, неизвестен, поэтому при многократной реализации однотипных процессов возникает статистическое распределение частоты исходов. Следовательно, R-преобразование может быть представлено распределением вероятности в ансамбле случайных событий, тогда такой ансамбль можно рассматривать как стохастическую систему. В связи с этим использование вероятностно-статистического аппарата для моделирования природных систем весьма перспективно.

В кибернетике при изучении технологических систем и их связей можно считать, что через "вход" элемент подвергается воздействию извне, а с помощью "выхода" сам воздействует на внешнюю среду. Такой несколько примитивный подход приемлем для речного водосбора или для любого элемента гидрогеологических систем. Для верхней гидродинамической зоны величина подземного стока во многом зависит от атмосферных осадков, а его воздействие на горные породы (внешнюю среду) может проявляться в виде механической или химической эрозии или через другие процессы, связанные с движением воды. Один из отправных и важных моментов при построении моделей гидрогеологических систем - выбор параметров внешней среды, представляющих непосредственный практический интерес, что связано с отсутствием тождественности внешних параметров для всех естественных систем.

Понятие системы предполагает существование взаимодействующих составных частей (элементов), которые принимаются в зависимости от задач исследования, и число их может быть неограниченным. В качестве элемента речного водосбора можно принять водосборы более низких порядков, а в дальнейшем - ландшафтные пояса, элементарные водоносные зоны и т.д.

Здесь мы опять сталкиваемся с важнейшим специфическим свойством

существования систем – эмерджентностью (это свойство характеризуется наличием у системы в целом таких качеств, которые не проявляются у каждого элемента в отдельности). Свойство получило название "парадокса Аристотеля": << Целое больше суммы своих частей >> . Простое на первый взгляд положение исходит из свойства детерминированности системы, но в приложении к гидрогеологическим системам выявление эмерджентных свойств во многом зависит от исследователя, выбирающего удобный способ представления моделирования.

Указанные свойства предполагают при изучении выявление иерархического строения (соподчиненности) и исследование совокупности связей между элементами и внешней средой. Примерами иерархической соподчиненности элементов гидрогеологических систем служат классификации подземных вод по типам скоплений, различные способы районирования гидрогеологических регионов; при анализе иерархической соподчиненности элементов исследователь сталкивается с проблемой классификации этих элементов, связанных в системе определенными логическими соотношениями, которые возможно моделировать. По А.Д. Арманду /1975/, модели природных систем – необходимое сырье, с которым имеют дело классификации (рис. 2).

Под моделью понимается совокупность теоретических и эмпирических представлений с целью получения и обобщения информации. Понятия "модель" и "система" тесно связаны между собой, но не тождественны. Системой является не всякая модель, а лишь формализованная, четко определяемая, создаваемая по определенным правилам познавательная конструкция /Жуков, 1978/.

Наиболее перспективными для большинства гидрогеологических задач могут явится статистические феноменологические модели, т.е. модели, конструируемые из особых элементов и позволяющие провести исследование форм проявления объекта как целого. Для процессов модели обычно называются функциональными, и при их построении приходится иметь дело с большим числом признаков, характеризующих эти процессы. Суть феноменологических моделей сводится к описанию процессов или пространственному распределению объектов через совокупность элементарных функций, которые выбираются с учетом априорной информации и гипотетических представлений об объектах. Назначение моделей заключается в том, что они должны отображать в некотором приближении природу рассматриваемых процессов. При построении статистических моделей приходится иметь дело также с многомерными выборками, которые в результате преобразований представляются в виде корреляционных (ковариационных) матриц и могут рассматриваться в виде структурной взаимосвязи элементов в системе.

Число элементов, их возможных состояний, количество связей между элементами в системе может быть чрезвычайно велико, особенно это относится к сложным природным системам. Известно, что связи между элементами системы и внешней средой могут рассматриваться в виде разнонаправленных потоков вещества, энергии и информации. Последняя разновидность потоков рассматривается в специфике воздействия одного элемента системы на другой, и в этом случае второй элемент – носитель информации о состоянии первого. В связи с этим возникла возможность отождествления сложных динамических систем с информационными /Арманд, 1975/. Информационные сигналы в таких системах реализуются в виде относительных изменений потоков вещества и энергии. По направленности связи в системах между элементами могут быть прямыми, обратными и направленными.

По мнению большинства исследователей, теория систем должна строиться на методах упрощения; один из таких методов для природных систем –

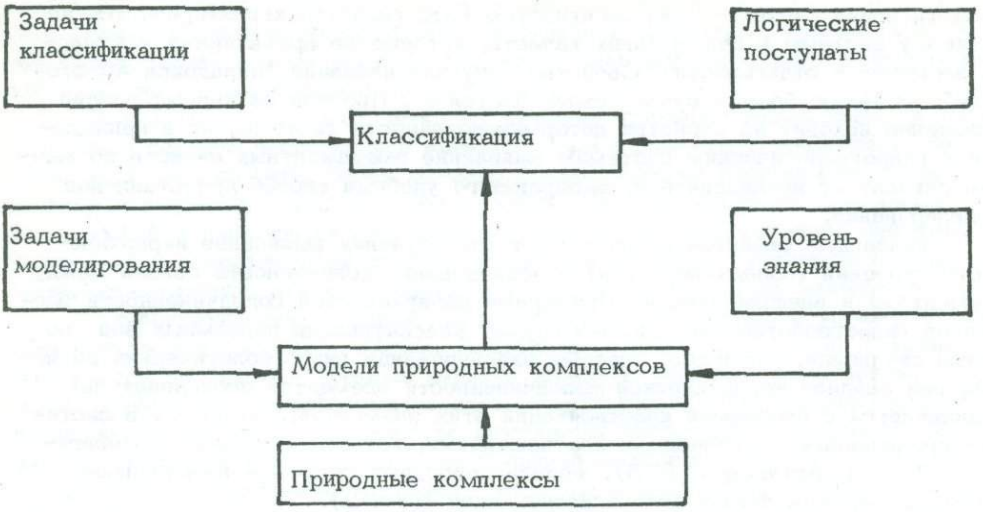


Рис. 2. *Место моделей в логической схеме построения классификаций природных комплексов* (по А.Д. Арманду /1975/).

объединение элементов по комплексу признаков в относительно независимые функциональные подсистемы и последующее моделирование происходящих процессов. *Подобные операции могут рассматриваться как классификация объектов* в многомерном пространстве или районирование с выделением различных уровней *иерархии подсистем, которые могут разрабатываться с различной степенью* детальности.

Разграничение системы с окружающей средой довольно условно, так как большую и сложную систему можно рассматривать как среду по отношению к *образующим ее меньшим подсистемам. Примером такой системы при изуче*нии формирования подземных вод может служить Монголо-Байкальский *орогенический пояс, состоящий из структурных подразделений более мелких иерар*хических порядков /Писарский, Хаустов, 1982/. Предлагаемая схема районирования *подземного стока по условиям формирования* (*табл.* 1) *предусматри*вает расчленение Монголо-Байкальского орогена на шесть таксономических *единиц от горно-складчатой системы до элементарной единицы - обводненной* зоны. Последнюю, в свою очередь, можно разбить на еще более мелкие структурные *единицы.* В *работах* Е.В. *Пиннекера* /1977/, К.П. *Караванова* /1977, 1980/, И.К. Гавич /1980/ в качестве наименьшей таксономической единицы *выделяется элементарная водоносная пора для пористых сред и элементарная* водоносная трещина для скальных пород. Однако поры и трещины в коллекторах *зачастую могут быть не связаны между собой, поэтому за низшее* структурное подразделение в предлагаемой схеме нами принята обводненная зона.

Основным критерием выделения наиболее крупной таксономической единицы - провинции являются геолого-структурные особенности региона. *Именно они в первую очередь определяют размеры и границы системы,* конструктивные черты распределения атмосферных осадков и последующие процессы их реализации в подземные воды. *Эти же условия определяют такие* важные свойства системы, как целостность, иерархичность, гетерогенность, эмерджентность и другие, *а также взаимодействие между подсистемами* различного иерархического уровня.

Провинция в зависимости от предыдущего критерия и времени тектони-

Таблица 1

Схема районирования горно-складчатых областей по условиям формирования подземного стока на примере Монголо-Байкальского орогена /Писарский, Хаустов, 1982/

Порядок (ранг)	Таксономическое подразделение	Система подземного стока	Критерии выделения	Индекс	Структурное подразделение	Территориальная единица
1	2	3	4	5	6	7
I	Провинция	Водоносная система	Геологическое строение региона	-	Горно-складчатая система	Монголо-Байкальский ороген
II	Область	Часть водоносной системы	Время и условия тектонической активизации	K_1	Кайнозойская активизация в условиях сжатия	Хангайское и Хэнтэй-Даурское неотектонические поднятия
				K_2	Кайнозойская активизация в условиях растяжения	Байкальский рифт
				M	Мезозойская активизация	Селенгино-Витимский прогиб
III	Округ	Группа бассейнов подземного стока	Геоморфологические и климатические условия	А	Горные хребты: в зоне избыточного и достаточного увлажнения;	Западно-Баргузинский, Хамар-Дабанский и др.
					в зоне недостаточного увлажнения	Орхонский, Северо-Хангайский и др.
				Б	Сочетание горных хребтов и межгорных впадин: в зоне избыточного и	Верхнеангарский, Баргузинский и др.

Продолжение табл. 1

1	2	3	4	5	6	7
					достаточного увлажнения;	
					в зоне недостаточного увлажнения	Джидино-Оронгойский, Хилокский, Верхнеселенгинский и др.
IV	Район	Бассейн подземного стока	Коллекторские свойства горных пород	1	Бассейны в карбонатных и интенсивно трещиноватых алюмосиликатных породах, рыхлых отложениях большой мощности	Слюдянско-Похабинский, Мысовой, Усть-Селенгинский и др.
				2	Бассейны в алюмосиликатных слабо трещиноватых породах, с рыхлыми отложениями малой мощности	Половинский, Рельский, Курбинский и др.
V	Подрайон	Пояс подземного стока	Типы ландшафта и характер изменчивости подземного стока	П	Приводораздельный пояс	Площади отсутствия или сезонного характера подземного стока в большинстве бассейнов
				Ск	Склоновый пояс	Площади постоянного подземного стока с высокой внутригодовой изменчивостью в

Окончание табл. 1

1	2	3	4	5	6	7
						большинстве бассейнов
		Д	Долинный пояс			Площади постоянного подземного стока с высокой зарегулированностью в большинстве бассейнов
VI	Участок	Часть пояса подземного стока	Различия геофильтрационной среды	-	Обводненная зона	Во всех бассейнах

ческой активизации разбивается на три более мелкие структурные единицы – области. На этом уровне таксономии наблюдается резкое различие как основных показателей подземного стока, так и условий, под влиянием которых он формируется (табл. 2). Очень высоки значения осредненных показателей подземного стока для области кайнозойской активизации – Байкальского рифта. Тектонически активная деятельность, особенно в неоген-четвертичное время, предопределила сочетание высокой степени увлажненности со зрелыми развитыми элементами рельефа и высокой степенью раздробленности горных пород. Это обеспечивает высокие коллекторские свойства водовмещающей среды и интенсивную инфильтрацию атмосферных осадков в подземные воды. В отдельных речных бассейнах значение модуля подземного стока достигает величин 15-17 л/(с·км^2) /Естественные ресурсы..., 1976/.

Хангайское и Хэнтэй-Даурское поднятия также испытывают активные тектонические движения, но развитие их еще не достигло стадии "зрелости", свойственной Байкальской рифтовой зоне. Для первых двух областей, так же как и для Байкальской, свойственны активная гидротермальная деятельность, проявляющаяся в зонах молодых сейсмоактивных разломов, и многочисленные признаки недавнего вулканизма /Писарский, 1982/. В силу особенностей географического положения этих горно-складчатых сооружений им свойственна меньшая величина увлажнения. Приведенные конструктивные особенности данных поднятий определяют несопоставимо меньшую величину осредненных показателей подземного стока (см. табл. 2).

Для Селенгино-Витимского прогиба (термин по Н.А. Флоренсову /1960/) свойственны менее благоприятные по сравнению с Байкальской рифтовой зоной условия формирования подземного стока, хотя территориальная близость этих регионов очевидна. Защищенность Селенгино-Витимского прогиба Байкальской рифтовой зоной от влагоносных ветров способствует меньшему увлажнению территории, а относительный текто-

Таблица 2

Осредненные показатели подземного стока в различных областях Монголо-

Активизация	Область	Тип округов подземного стока
Современная	Хангайское поднятие	В горных хребтах
	Хэнтэй-Даурское поднятие	То же
Кайнозойская	Байкальский рифт	"
		При сочетании горных хребтов и межгорных впадин
Мезозойская	Селенгино-Витимский прогиб	То же

нический покой в кайнозойскую и современную эпохи (показатели в 4-5 раз ниже по сравнению с рифтом) создает менее благоприятные емкостные свойства водовмещающих горных пород. Эти факты сказываются на значениях основных показателей подземного стока (в 5-10 раз ниже, чем в рифте) и косвенных (примерно ниже в 2 раза).

Следующее таксономическое подразделение - округ подземного стока объединяет группу речных бассейнов на основе сходства геоморфологических и климатических особенностей региона. Выделяются два типа округов: в гидрогеологических массивах и при сочетании горных хребтов с межгорными впадинами. Различия условий формирования подземного стока в этих типах округов заключаются в его количественных параметрах, характере временной изменчивости и степени активности его взаимосвязи с поверхностным стоком.

В областях современной активизации известны только округа горных хребтов, а в Селенгино-Витимском прогибе - округа, где сочетаются горные хребты и межгорные впадины.

В пределах Байкальского рифта развиты оба типа округов. Для первого типа характерны более высокое увлажнение, интенсивные уклоны местности, отсутствие мощного чехла рыхлых отложений и по сравнению со вторым типом меньшие расчлененность рельефа и плотность тектонических нарушений. Перечисленные условия формируют в первом типе преимущественно нисходящий (гидравлически не связанный) подземный сток в реки, характеризующийся высокой внутригодовой и многолетней изменчивостью по сравнению с группами речных бассейнов, имеющих межгорные впадины. Если для рек, формирующих подземный сток гидрогеологических массивов, величина модуля в среднем 4-6 л/(с·км^2), то для групп бассейнов второго типа показатели почти на порядок ниже, особенно для впадин забайкальского типа.

Байкальского орогена /Писарский, Хаустов, 1982/

Средние показатели подземного стока				
Модуль, $л/(с \cdot км^2)$	Коэффициент подземного стока, % / Коэффициент подземного питания, %	Общее увлажнение, мм / Уклон рельефа, ‰	Комплексный показатель расчленения рельефа, м/км / Плотность тектонических нарушений, $км^2/км^2$	Относительный градиент скорости вертикальных тектонических движений, $n \cdot 10^{-9}$ см/год
0,2	1/11	400/60	18/-	-
0,8	4/12	350/140	212/17	-
4,5	17/48	980/220	420/28	1,5
5,2	18/40	900/190	560/35	1,2
0,5	3/20	460/120	238/18	0,3

4706

Основная таксономическая единица, с которой в дальнейшем предполагается проводить построения статистических моделей формирования ресурсов подземных вод, – район подземного стока, которому в горно-складчатых областях территориально соответствует площадь речного водосбора. При этом принимается, что границы поверхностного водосбора совпадают с подземными, за исключением площадей, где широкое развитие имеют карбонатные породы. Такие районы на предлагаемой схеме выделяются в отдельную категорию и на территории Байкальской горно-складчатой области имеют ограниченное распространение (см. табл. 1). Доминирующим критерием, определяющим здесь условия формирования подземного стока, выступают коллекторские свойства горных пород как для зоны аэрации, так и для зоны насыщения. Эти свойства зависят в первую очередь от литологического состава пород, степени их трещиноватости и тектонической раздробленности.

Для межгорных впадин байкальского типа, выполненных мощной толщей рыхлых кайнозойских осадков, большое значение имеет пористость пород. Во впадинах забайкальского типа, где мезозойские породы метаморфизованы в процессе геологического развития, решающая роль переходит к литогенетической и тектонической трещиноватости. В качестве примера можно сравнить отдельные речные бассейны Ангаро-Баргузинской группы, находящейся в зоне избыточного увлажнения, и Джидино-Удинской – в зоне недостаточного увлажнения. Величины модулей подземного стока для рек первой группы изменяются от 2 до 10 $л/(с \cdot км^2)$, а для второй – от 0,08 до 2,0 $л/(с \cdot км^2)$.

В пределах речных бассейнов в качестве подрайонов выделяются пояса подземного стока, в которых его формирование находится под влиянием ландшафтных условий. В следующей главе, а также в других работах /Писарский

и др., 1975; Писарский, Хаустов, 1979/ более детально рассмотрены основные принципы поясной структуры формирования подземного стока в горно-складчатых областях. Характеристики подземного стока и основные его параметры довольно четко отличаются для разнообразных типов поясов: приводораздельного, склонового и долинного. Пояс подземного стока является как ранговым в пределах речного бассейна, так и внеранговым таксономическим подразделением и может быть выделен даже в пределах провинций горно-складчатых сооружений.

Наиболее мелкая иерархическая единица в предлагаемой схеме - участок подземного стока, под которым понимается обводненная или водоносная зона. Это может быть водоносный комплекс, горизонт, пласт, блок, разлом или серия горных пород с относительным единством геофильтрационной среды. При переходе от одних таксономических единиц к другим, наряду с указанием основных критериев выделения, подразумевается сохранение влияния ранее отмеченных условий, т.е. в направлении от провинции к участку идет постоянное увеличение числа критериев выделения с сохранением их иерархической значимости.

Формализованное описание системы наиболее эффективно при использовании макроподхода, сущность которого состоит в изучении отдельных структурных элементов системы, их "входов" и "выходов", функций, а также внутрисистемных связей. При таком подходе систему расчленяют на структурные элементы разных масштабов и уровней организации; выясняют роль каждого структурного элемента в преобразовании и классифицируют их сообразно этому преобразованию; изучают особенности структуры, обеспечивающие устойчивость системы и ее способность к саморегулированию; для каждого элемента или их класса находят алгоритм преобразования потоков энергии, вещества, информации; выявляют внутрисистемные связи элементов, их характер и роль; находят синтетический алгоритм системы /Булавко, 1971; Оноприенко, 1974/. В работе В.В. Дружинина и Д.С. Конторова /1976/ описание всякой новой системы рекомендуется начинать с характеристики трех точек зрения: функциональной, морфологической, информационной.

Функциональное описание необходимо для осознания важности места и оценки отношения к другим системам (подразумевается прежде всего выявление внешних связей систем). Обращаясь к табл. 1, можно видеть, что функциональное описание проводится на любом уровне таксономии, т.е. графа, определяющая ранг системы, прежде всего указывает на соподчиненность систем мелкого таксономического уровня более крупным (от участка к провинции). Например, для бассейна подземного стока выявляется наличие или отсутствие гидравлической связи с соседними речными бассейнами. В свою очередь, он разбивается на более мелкие таксономические единицы - пояса, и наоборот, совокупность речных бассейнов - среда для формирования более крупной таксономической единицы - округа. На этом уровне, наряду с фильтрационными свойствами горных пород, основные критерии выделения - геоморфологические и климатические условия.

Морфологическое описание должно давать представление о строении системы и так же, как и функциональное, может быть иерархично и многоуровенно. Каждое такое описание должно исходить из конкретно поставленной задачи. Общая сумма иерархии морфологического описания систем /Дружинин, Конторов, 1976/ приведена на рис. 3. По составу элементов системы могут быть гомогенными (однотипными), гетерогенными (разнотипными) и смешанными. В природных условиях чаще всего преобладают гетерогенные системы. Например, водоносный горизонт или водоносная зона для кристаллических пород могут быть представлены (по вещественному составу) твердым скелетом вмещающих пород, движущихся по системе пор или трещин водой и растворенными и свободными газами. Гетерогенность систем низкого уровня порождает гетерогенность высокого, и наоборот. Так, для артезианского бассейна степень гете-

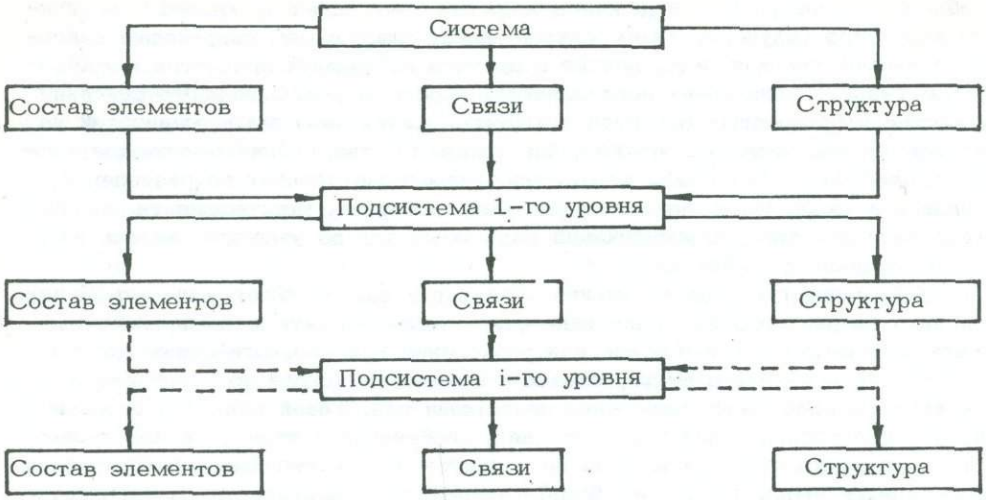


Рис. 3. Иерархия морфологического описания систем /Дружинин, Конторов, 1976/.

рогенности определяется числом слагающих его водоносных горизонтов, а совокупность гетерогенных артезианских бассейнов формирует гетерогенность области (систематика по Е.В. Пиннекеру /1977/).

Структурные свойства систем определяются характером и устойчивостью связей между элементами. По этому принципу выделяются системы многосвязные, иерархические, смешанные. Низкий уровень их структурных свойств определяется полной неизменностью. В гидрогеологических исследованиях чаще всего приходится иметь дело с высокоуровенными структурными свойствами и связями.

Информационное описание системы заключается в познании ее организации. Высокоуровенные системы при изучении предполагают на различных стадиях наращивание структуры (включение и выключение определенных элементов и связей). В качестве примера можно сослаться на табл. 1, где по мере перехода к более низким таксономическим единицам происходит увеличение числа элементов системы и их связей, т.е. наглядно демонстрируется один из основных принципов информационного описания - расчленение системы на элементы. Такие задачи обычно решаются с помощью методов типологии и классификации.

Необходимость представления гидрогеологических объектов как систем в виде связей и применение к решению прикладных задач с целью управления поверхностными и подземными водными ресурсами достаточно подробно изложены ранее /Умаров, 1982; Angewandte Hydrologie, 1980; Luckner, 1975; и др./.

1.2. МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ ТИПОЛОГИИ И КЛАССИФИКАЦИИ ГИДРОГЕОЛОГИЧЕСКИХ СИСТЕМ

Под типологией обычно понимается метод научного познания, базирующийся на расчленении изучаемой совокупности объектов в однотипные группы по определенным свойствам с помощью идеализированной модели. С помощью

подобных методов решается довольно широкий круг задач прикладного и научного характера (изучение взаимосвязей между признаками, сокращение описания изучаемой совокупности, анализ и прогноз изучаемых процессов и др.).

В гидрогеологических исследованиях проблема упорядоченного описания множества неоднородных объектов и изучение каких-либо закономерностей на основании таких множеств чрезвычайно остра. Подтверждением является большое количество наименований структурно-гидрогеологических подразделений, принципов и видов гидрогеологического районирования, многовариантность гидрогеологических карт, классификаций подземных вод по условиям залегания и другим различным свойствам.

Типологию гидрогеологических объектов в самом общем виде можно подразделить на две основные разновидности - теоретическую и эмпирическую. Первая подразумевает обобщение или разделение гидрогеологических систем на основе понятий о структуре, связях и элементах исходя из логического анализа закономерностей формирования подземных вод. Такой подход в геологических и гидрогеологических исследованиях, особенно при типизации водоносных систем, сложился сравнительно давно и является традиционным. Эмпирическая типологизация предполагает устойчивые сочетания совокупности свойств гидрогеологических объектов. Устойчивые сочетания совокупности свойств объектов, выраженные через определенные координаты (меры), находятся путем многократного перебора свойств. Такое направление в гидрогеологии возникло сравнительно недавно и весьма перспективно, поскольку с помощью этих процедур можно снять те неопределенности, о которых говорилось выше, особенно при всех видах гидрогеологического районирования.

Проблема районирования - одна из важнейших задач, стоящих перед современной гидрогеологией. В работе Е.В. Пиннекера /1977, с. 23/ отмечается, что «...в задачу гидрогеологического районирования входит систематизация пространственных закономерностей формирования подземных вод... ». В широком смысле под любым видом районирования (дополнительно вводятся координаты географического пространства) можно понимать классификацию объектов, на которую накладывается ряд жестких ограничений, и, следовательно, рассматривая классификацию как задачу гидрогеологического районирования, исследователь в значительной мере объективизирует поставленную задачу.

Н.И. Кондаков /1971, с. 214/ дает такое определение классификации « ...распределение предметов какого-либо рода на классы, согласно наиболее существенным признакам, присущим предметам данного рода и отличающим их от предметов других родов ». Под классом Н.И. Кондаков понимает совокупность объектов, имеющих один или несколько характеристических признаков. При таком определении класс можно отождествить с гидрогеологическим районом, областью, бассейном подземного стока, поясом и т.д. В этом определении классификации видится логическое соответствие определению районирования, данному Е.В. Пиннекером.

Районирование заключается в выборе классификационных признаков и проведении на карте границ в местах, где признаки изменяются /Основы..., 1980/. При районировании подземного стока для территории СССР главный принцип - структурно-гидролого-гидрогеологический. В качестве элементарной таксономической единицы принимался речной бассейн стока. Правомерность выделения такой элементарной единицы при районировании Б.И. Куделин и И.Ф. Фиделли /Подземный сток..., 1966; Куделин, Фиделли, 1970; Фиделли, 1980/ аргументируют единством структуры водного баланса этой единицы, что особенно ярко и наглядно проявляется в горно-складчатых областях. В.А. Всеволожский /1983/, рассматривая возможности определения границ гидрогеодинамических систем для платформенных структур, также приходит к

выводу о единственно обоснованном принципе с проведением границ по водоразделам поверхностного и подземного стока. При этом порядок системы определяется порядком базиса дренирования водоносных систем.

Следовательно, речной бассейн или его часть характеризуются совокупностью структурно-геологических, гидрогеологических, морфометрических, гидрометеорологических и других признаков или их полями, результатом взаимодействия которых является подземный сток. Рассматривая речной бассейн с набором признаков как вектор или точку в многомерном пространстве, при условии их естественных группировок, мы можем применять аппарат многомерного анализа, в частности алгоритмы классификации для целей районирования. Эта операция требует дополнительного введения координат пространства и может быть реализована вручную и автоматически с помощью ЭВМ.

Различие между группами (классами) этих точек предопределяется неоднородностью и масштабностью комплекса природных процессов, оказывающих влияние на формирование естественных ресурсов подземных вод. Нахождение всевозможных сочетаний признаков, характеризующих изменчивость количественных показателей подземного стока, дает возможность реально оценить масштабы воздействия природных условий. Такой подход позволяет рассматривать гидрогеологическое районирование как одну из классификационных задач.

В гидрогеологии, как и в других естественных науках, районирование - метод исследования закономерностей, свойственный явлениям и процессам различного масштаба, и может осуществляться с помощью теоретической и эмпирической типологий. Тот и другой принцип не лишены недостатков. Если первый (в той или иной степени) несет определенную долю субъективизма, то второй может быть во многом формальным, а отсюда, как следствие, его неадекватность природным условиям. Наиболее разумный подход к районированию классификации природных объектов - сочетание этих двух подходов.

Рассматривая классификацию как основу той или иной модели, следует представлять, что она должна не только объяснять информацию, но и отражать реальную действительность природных процессов, являться эффективным аппаратом исследования.

Под функциями классификации понимается ее свойство удовлетворять какую-либо потребность. В работе В.Л. Кожары /1982/ с этих позиций классификации подразделяются на две разновидности:

классификации информирования, которые призваны удовлетворять потребность в имеющихся сведениях о познаваемой действительности и, следовательно, должны нести функции хранения, поиска, обобщения и передачи информации;

классификации прогнозирования, которые базируются на свойствах выборочных объектов, должны удовлетворять прежде всего потребность в познании того или иного процесса при отсутствии эмпирических сведений.

Процедура классификации природных объектов должна удовлетворять следующим детерминистическим принципам /Кондаков, 1971; Николаев, 1976; Хаустов, 1982б; Sneath, Sokal, 1973/.

1. Число классов не определяется заранее, оно выявляется в процессе решения поставленной задачи (при априорном задании числа классов классификация не может быть объективной, так как в самом начале на решение задачи накладываются ограничения).

2. Число классов должно быть существенно меньше числа совокупности классифицируемых объектов.

3. Выделенный класс не должен быть пустым (т.е. не содержащим ни одного реального объекта, иначе теряется целевое назначение классификации).

4. При классификации объектов необходимо использовать одни и те же признаки (основания). (Это, однако, не исключает возможности использования

на различных стадиях классификации дополнительных признаков, поскольку в ней будет зафиксировано большее число свойств и отношений объектов.)

5. Ни один из объектов классификации не может одновременно входить в разные классы (т.е. классы должны быть дизъюнктивные, а их образы непересекающиеся).

6. Параметры совокупности объектов одного класса должны существенно отличаться от таковых другого.

Задачи районирования, как и задачи классификации, предусматривают несколько этапов: выбор исходной системы признаков исследуемых объектов, преобразование признаков в удобную для классификации систему, принятие решения о числе классов при возникновении такой необходимости, отнесение отдельных объектов к тому или иному классу, содержательная интерпретация полученных классов (районов), объективная оценка степени логичности и достоверности предложенной классификации. При классификации природных объектов могут возникнуть варианты, при которых совокупность признаков не обладает дискриминационными качествами. Вследствие этого не происходит четкого разделения объектов на компактные множества и между ними не существует выраженного пространства для уверенного проведения границ множеств.

Эта проблематичность во многом может быть снята обращением к постановке задач исследования (блок II), а также к исходным данным (см. рис. 1, блоки III, IV). Иными словами, общее представление о характере процессов в системе и ее элементах (классах) у исследователя должно быть сформулировано при постановке задачи и выборе путей ее решения. Не исключено, что после реализации одного или нескольких алгоритмов эти представления могут быть коренным образом пересмотрены. Проверка, отбор и сравнение работы алгоритмов могут проводиться на основе использования искусственных эталонов с заданными статистическими свойствами /Гавришин, 1980а; Gabriel, 1968/. Весьма полезно при выборе алгоритмов классификации обсуждение их со специалистами по вычислительной технике, поскольку почти всегда возникают трудности при реализации этих алгоритмов на ЭВМ.

Возможные несоответствия получаемых результатов при использовании различных алгоритмов многомерной группировки на одной и той же выборке не являются недостатком их (алгоритмов) работы, а скорее наоборот, различные методы открывают дополнительные представления о типе объектов и процессах /Типология..., 1982/.

В современной литературе немало дискуссий о необходимом числе объектов и признаков для целей классификации, при этом считается, что, чем большее число объектов используется, тем объективнее получается классификация. Для природных процессов этот вопрос также немаловажен, и, по-видимому, решающее слово остается здесь за исследователем. Главный принцип, который должен соблюдаться, - это равномерный охват территории, если классификация рассматривается как районирование, и наличие носителей информации во всех ландшафтных зонах региона. Сколько бы объектов ни было положено в основу классификации, мы фактически будем иметь дело с некоторой выборкой из генеральной совокупности.

С нашей точки зрения, наиболее важным и ответственным этапом является отбор информативных признаков или формирование признакового пространства. Здесь уместно привести высказывание А.А. Френкеля (цитируется по /Типология..., 1982, с. 18/): « Математическая статистика оперирует не с множеством объектов, а с множеством исходов наблюдений, то есть значений исследуемого показателя...». В нашем случае это можно отнести к набору признаков.

Оперируя с такими природными объектами, как речной бассейн, и с таким сложным процессом, как формирование подземных вод, мы чаще всего

сталкиваемся с такой проблемой эмпирической типологии, когда нельзя заведомо точно охарактеризовать типы изучаемых объектов и нет уверенности, что это именно те признаки, которые играют ведущую роль в изучаемых нами процессах. Неизвестно, достаточно ли надежно эти признаки позволяют отнести изучаемые нами объекты к тому или иному классу и, наконец, с какими признаками объекты можно считать однотипными.

Как было показано В.Л. Кожарой /1979/, при "естественной" классификации природных объектов с числом оснований (признаков) от 170 до 30 при различном количестве объектов информация насыщается. Иными словами, начиная с некоторого сравнительно большого объема информативность признаков для целей классификации увеличивается пренебрежимо мало. В этой же работе классификации строятся на генетически однородных объектах, в результате чего получаются близкие по генезису классы, характеризующиеся малым отличием параметров. Такие задачи наиболее трудны, поскольку в типологических координатах межклассовые расстояния чрезвычайно малы и не всегда определены, а границы между ними размыты. Тем не менее эти задачи при классификации природных объектов встречаются очень часто и имеют большой практический и теоретический интерес. Примером служит опыт классификации речных бассейнов по условиям формирования подземного стока с помощью Q-схемы факторного анализа в предлагаемой работе. Подавляющее большинство факторных нагрузок на первый признак, при его вкладе в общую дисперсию 95%, соответствует величинам от 0,96 до 1,0. Это подтверждается и корреляционной матрицей между 80 объектами по 14 признакам. Значения коэффициентов корреляции изменяются от 0,86 до 1,0.

В работе В.Л. Кожары /1979/ также приводятся данные по устойчивости классификаций для различных естественных процессов. Для климатов оптимальная величина оснований (признаков) оценивается 40-50 значениями, для классификации физико-химических свойств природных вод Алданского нагорья исчерпывается 15 признаками. Этим же автором рекомендуется для формализации построения "естественных" классификаций брать легко и точно измеряемые характеристики, не слишком коррелируемые в случайном сочетании.

Из приведенных положений можно сделать весьма важный вывод: для построения классификаций природных систем разной степени динамичности требуется различное число признаков; чем выше степень динамичности системы, тем большее число признаков должно быть положено в основу классификации систем или процессов. Все рассматриваемые признаки могут быть статистически однородными и неоднородными. В качестве критерия однородности признаков, используемых для классификации, некоторые авторы рекомендуют принимать коэффициенты их вариации, не превышающие 0,33 /Сиськов, 1975/.

С таким жестким условием, по нашему мнению, трудно согласиться, поскольку предполагается, что наибольшую информацию, особенно о высокодинамичных процессах, несут признаки с высокой степенью вариабельности. Эти признаки позволяют более четко ограничить межклассовые расстояния и сместить центры тяжести классов в координатных плоскостях в ту или иную сторону. Такое смещение может быть достигнуто при помощи последовательной замены или перебора совокупности признаков с использованием методов многомерных группировок.

В основу любого метода классификации положен принцип разделения исходной совокупности объектов на классы, состоящие в какой-то степени из близких по значению признаковому пространству объектов, поэтому необходимо введение понятия "однородность класса". Для оценки этой характеристики используется достаточно большое число статистических критериев, которые не все по своей разрешающей способности одинаковы, и отсюда вытекает многовариантность классификационных схем на одних и тех же выборках в зависи-

мости от применяемых алгоритмов. Для преодоления неопределенности возможно использование искусственных эталонов, на базе которых вырабатываются достаточно четкие критерии однородности классов.

Иногда при классификации природных объектов не удается получить достаточного разобщения исходных данных, а межклассовые расстояния имеют размытые, не четко выраженные границы. В этом случае можно прийти к выводу об отсутствии дифференциации природной среды. Здесь могут быть две серьезные причины: либо используемый алгоритм имеет недостаточно гибкую метрику, на основании которой происходит объединение объектов в классы, либо исходная совокупность признаков не совсем полно отражает индивидуальные черты объектов, используемых для классификации.

Под метриками обычно понимается функция близости (или расстояния) между двумя рассматриваемыми объектами, выраженная в виде математического закона. Введением метрик, как правило, определяется однородность класса, отсюда вытекает важность выбора вида метрик при разработке алгоритмов классификации. Применению последних должны предшествовать оценка статистических свойств совокупности признаков: асимметрии, вариации, эксцессов; расчеты критериев проверки соответствия эмпирического распределения многомерному нормальному и другие элементарные статистические процедуры.

Обычно на этом уровне системного анализа мы вплотную сталкиваемся с понятием оценки адекватности классификационных моделей (см. рис. 1, блок VII). Конструирование моделей (блок VI) не исключает такой возможности, как использование на отдельных этапах классификации различных алгоритмов и анализ их работоспособности. Необходимо подобрать такой формальный алгоритм классификации, чтобы каждый класс с наибольшей вероятностью формировал однотипные объекты. При этом вряд ли перспективны алгоритмы, основанные на методах полного перебора, когда исследуются возможные комбинации объектов и их объединение с помощью определенных критериев.

Выбор конечного шага итерации многомерных классификаций – также весьма важный момент, который является необходимым звеном при оценке адекватности таксономических моделей. Степень адекватности для природных систем, да и не только природных, определяется уровнем прогностичности (способностью выделенных классов давать максимально возможную информацию о тех процессах, которые происходят в выделенных природных пространствах). Конечный итог итерационных процедур при многомерных классификациях должен дать ответ: до какого уровня дробить исходную совокупность объектов. Прежде всего конечный шаг итерации должен задаваться самим исследователем в зависимости от поставленных задач. Можно разбивать исходную совокупность до тех пор, пока каждый объект многомерных наблюдений не будет считаться классом. В сущности, содержательная интерпретация классов – конечный итог классификации.

Природным процессам свойственна масштабность – воздействие процессов различного генетического характера на изучаемое нами явление. Поскольку природные процессы можно рассматривать как результат взаимодействия комплекса физических полей, характеристики которых обладают пространственно-временной изменчивостью, изучаемые нами поля на различном уровне метризации, а следовательно, и таксономии будут иметь микро-, мета- и макронеоднородности, т.е. приобретать свойство масштабности. Эта масштабность создается суперпозицией определенного набора полей, которые должны быть идентифицированы при помощи интерпретации классов на различных иерархических уровнях классификации систем.

Интерпретацию классов можно начать с анализа признаков, по которым была проведена классификация, и отыскания интересующих закономерностей или процессов внутри этих классов. Вначале строится логическая модель про-

цесса и выделяются признаки причины и признаки проявления, т.е. изучаются причинно-следственная связь и ее различия в каждом классе (районе). На этом же этапе необходимы анализ соответствия работы алгоритмов классификации и для облегчения интерпретации классов - применение других видов моделей, например построение и анализ пространственно-временных корреляционных и взаимокорреляционных функций, оценка вариабельности и других статистик признаков внутри классов, построение регрессионных или факторных моделей и иных разновидностей статистического анализа. Поскольку любая классификация (а районирование вне всякого сомнения) во многом призвана решать прогностические задачи, то очень важными представляются построение и анализ регрессионных моделей признаков причин и признаков проявления по общей выборке и по каждому выделенному классу.

Методология целей и способов интерпретации выделенных классов зависит прежде всего от постановки проблем и задач исследования. Если исходный набор признаков, на базе которых строится классификация, не приводит к положительным результатам, его нужно существенно изменить или дополнить новыми - набор признаков должен быть не только достаточным, но и необходимым. Как правило, наибольший эффект классификации объектов достигается при использовании максимального числа признаков, характеризующих данный объект, но в этом случае все признаки в равной мере являются "информативными". Возникает проблема их отбора для целей классификации. Изменение количества признаков может существенным образом повлиять на число и параметры классов, а также на распределение объектов и их принадлежность. Здесь уместно привести "теорему о гадком утенке", согласно которой нельзя выделить классы, объединяющие сходные объекты, если считать все признаки в равной степени "информативными" /Ватанабе и др., 1969/. В связи с этим необходимы преобразования исходной информации с целью выделения главных (основных) признаков совокупности классифицируемых объектов и избавление от мелких второстепенных деталей (шумов).

Одно из необходимых условий при этих процедурах - незначительные *потери исходной информации*, что может быть решено с помощью линейного преобразования исходной информации в признаковом пространстве (метод главных компонент, факторный анализ сингулярное разложение, канонические корреляции). Идея методов состоит в отыскании некоррелированных линейных комбинаций признаков, или в геометрическом представлении как переход к новой системе координат, на которые падают максимальные дисперсии признаков.

В результате преобразования получается система взаимно независимых признаков (факторов), что в значительной мере упрощает процесс классификации и позволяет определить соотношение между признаками исходной выборки, их количеством и степенью информативности. Дальнейшая классификация может осуществляться по полученным в результате преобразований новым "синтетическим" характеристикам.

Под адекватностью моделей прежде всего понимается их соответствие условиям, реально существующим в природе. На этом этапе системного анализа неизбежно приходится обращаться к процедуре сравнения работы алгоритмов, используемых для построения многомерных моделей. Для корреляционных и регрессионных моделей адекватность довольно просто выявляется с помощью моделей типа процесс - отклик или процесс - результат. Эти модели здесь понимаются по У. Крамбейну и др. /1973/. Принимая в левой части уравнения регрессии за переменную интересующий нас процесс, рассматриваемый как отклик, который связан с одной или более переменными, в правой можно путем увеличения их числа добиться желаемой степени адекватности, сравнивая получаемые данные с эмпирическими. При этом довольно несложно выявить ошибку расчетов по предлагаемой модели в зависимости от числа предикторов.

Намного труднее обстоит дело с многомерными классификациями. До настоящего времени нет сколько-нибудь четких критериев отражения степени адекватности автоматических классификаций для природных систем. Чаще всего критерии качества характеризуют лишь методологическую сторону процедуры или проверку работы алгоритмов на искусственных эталонах. Эти моменты - очень важные и неизбежные этапы построения классификаций, но для природных систем вопрос адекватности классификаций до сих пор остается открытым. Весьма интересное направление в этой области - метод экспертных оценок, на основании которых могут быть выработаны различные критерии адекватности.

Сказанное не снижает ценности многомерных моделей как мощного метода познания природных систем. В случае получения различных классификаций при использовании работы нескольких алгоритмов на одних и тех же выборках предпочтение отдается тому варианту, который наилучшим образом интерпретируется и соответствует принятой гипотезе, лежащей в основе представлений об интересующем процессе. В этом случае каждая классификация - средство к достижению поставленной цели. Главный принцип, который здесь должен сохраняться, - это комплексирование процедуры типологизации с другими методами статистической обработки данных, начиная с вычислений простых статистик для каждого класса и всей совокупности, вплоть до методов факторного анализа. Эти методы позволят в значительной мере избежать того формализма, который присущ большинству алгоритмов.

Глава 2

МЕТОДИКА РАСЧЕТА ПРИЗНАКОВОГО ПРОСТРАНСТВА МНОГОМЕРНЫХ МОДЕЛЕЙ ФОРМИРОВАНИЯ РЕСУРСОВ ПОДЗЕМНЫХ ВОД

Успех практического использования многомерных моделей, конечные результаты и их интерпретация прежде всего зависят от особенностей совокупности признаков, используемых для построения этих моделей. Формирование признакового пространства приобретает особую роль в задачах таксономии, проблемам которой применительно к образованию естественных ресурсов подземных вод в горно-складчатых областях посвящена основная часть монографии.

Формирование подземного стока происходит стихийно, под влиянием множества процессов и условий. Свойственные ему закономерности проявляются статистически с множеством частных отклонений, например в виде пространственно-временной изменчивости от некоторой тенденции - тренда. Эта изменчивость порождается вариацией природных процессов, воздействующих на подземный сток в той или иной мере. Там, где такие связи проявляются более сильно, подземный сток наследует их черты, и возможен его прогноз по косвенным признакам, например по величине увлажнения региона, расчлененности рельефа, геолого-структурным данным и др. В противном случае необходим поиск новых признаков, либо переход к новым шкалам измерений, либо существенное уточнение старых признаков. Полностью формализовать описание природных процессов нельзя, можно лишь предложить ряд операций, позволяющих облегчить исходную информационную систему признаков с целью отбора наиболее существенных с нашей точки зрения.

При построении моделей мы неизбежно сталкиваемся с проблемами оценки достаточности и необходимости некоторой совокупности исходных признаков. Этот процесс неразрывно связан с классификацией и может быть условно разбит на два этапа: выбор исходного множества признаков с тем, чтобы они достаточно полно характеризовали объект или природный процесс, и конструирование по этой системе признаков их новой совокупности, несущей наибольшую информационную емкость, которая достаточно полно отражает классификационные свойства процессов или объектов. Иными словами, необходимо провести с помощью какого-либо метода качественную отбраковку признаков для последующего построения многомерных моделей. Такие итерации могут повторяться неоднократно с использованием различных методов и приемов, например факторного, компонентного или регрессионного анализов, последовательного введения случайных чисел и др. Возможно также объединение групп в новые синтетические переменные или оперирование различными мерами сходства, например построение корреляционных моделей и дальнейшее их использование для целей классификации.

Поскольку последние задачи относятся к классу типологических, то здесь в первую очередь встает вопрос о двойственном противоречии, суть которого в следующем: пока неизвестно разбиение исходного множества на подмножества, нельзя оценить степень информативности связей, а следовательно, и правильно построить классификацию, поскольку классификации должны строиться на правильно сконструированной системе признаков. Преодоление этого противоречия возможно с помощью итеративных последовательных процедур, цель которых - выявление все более надежных связей между признаками.

Некоторые авторы, например В. Плюта /1980/, считают, что сильно скоррелированные признаки не дают существенного прироста информации для построения многомерных классификаций, а отбор информативных признаков должен строиться на учете слабых корреляций между собой в системе, но эти признаки должны сильно коррелироваться с неучтенными. С таким доводом трудно полностью согласиться. Для природных систем, в частности гидрогеологических, поиск неучтенных признаков и оценка их связей с учтенными информативными - процедура весьма условная, поскольку наши знания о системах и происходящих в них процессах не столь полны, как в технических, экономических или социальных системах. Каждый признак, полученный при гидрогеологических исследованиях и раскрывающий хотя бы малую часть свойств объектов, представляет большой научный и практический интерес.

При использовании достаточно сильно скоррелированных признаков может возникнуть ситуация, когда признаки, имеющие сильные корреляционные связи между собой для всего множества, в подмножествах имеют значительно меньшие коэффициенты или становятся вообще независимыми, а само по себе снижение ценности признаков в итерационных классификациях - довольно распространенное явление. В этом случае необходим переход к многоуровенным классификациям, заключающимся в последовательном дополнении или сокращении признакового пространства. Для решения классификационных задач проблема формирования признакового пространства чаще всего сводится к его преобразованию и четкому пониманию целей классификации.

Проблема достаточности признакового пространства может быть решена поиском их оптимума при заданной таксономической структуре совокупности объектов. Зная число классов и решающую функцию, с помощью которой классифицируются объекты, можно определить необходимый минимум признаков, позволяющий с заданной точностью отличать таксоны друг от друга.

В самом общем виде все признаки, с помощью которых можно охарактеризовать объекты или процессы, подразделяют на следующие группы: дефиниционные, являющиеся частью объекта или процесса; являющиеся их причи-

ной или условием, их следствием, их коррелятом /Типология..., 1982/. С позиции изучения формирования подземных вод, придерживаясь этой типизации, к первой группе можно отнести геолого-структурные признаки; вторую составят метеорологические признаки (в основном это осадки и испарение); третью сформируют основные параметры подземного стока; четвертую - морфометрические и геоморфологические показатели. Следует отметить, что такое подразделение весьма условно и нуждается в дальнейшей корректировке, но уже на этой стадии оно может быть полезно при анализе генетических группировок признаков и полноты характеристики объектов. При классификации гидрогеологических систем приходится иметь дело с различными признаками, которые не всегда удается выразить через метрические значения, точность же измеренных признаков не всегда равнозначна и оставляет желать лучшего. Необходимо помнить, что никакая сверхЭВМ и супермодель не смогут улучшить исходную информацию, они призваны расширять представления об объектах и методах их исследований.

Широкое использование основных и косвенных признаков позволяет объективно оценить пространственно-временную изменчивость процессов формирования подземного стока, провести их классификацию, определить пространственно воздействие комплекса признаков и их роль для различных условий образования ресурсов подземных вод.

На данном этапе исследования основные признаки процессов формирования подземного стока в бассейне оз. Байкал, как и в других горно-складчатых областях, можно условно разбить на следующие группы: 1 - основные параметры подземного стока; 2 - геолого-структурные, 3 - морфометрические и геоморфологические, 4 - гидрометеорологические признаки; 5 - характеристики поясности подземного стока.

2.1. ОСНОВНЫЕ ПАРАМЕТРЫ ПОДЗЕМНОГО СТОКА

Основными параметрами подземного стока являются модуль ($M_{п.с}$); коэффициент подземного питания ($K_{п.п}$), представляющий отношение величины подземного стока к величине общего речного; коэффициент подземного стока ($K_{п.с}$), рассчитываемый как отношение слоя подземного стока к слою атмосферных осадков или величине увлажнения. Методика оценки величин подземного притока в реки Байкальского региона с применением гидрохимических данных и результаты этой оценки для большинства речных бассейнов определены /Писарский, Хаустов, 1973; Федоров, Хаустов, 1973; Естественные ресурсы..., 1976/ и здесь детально не приводятся.

Метод основан на выявлении комплекса связей между изменением минерализации воды в реках бассейна оз. Байкал в зависимости от их расходов. В качестве маркирующего ингредиента выбран гидрокарбонат-ион, при этом весьма важным является обоснование получения его средних характеристик различной обеспеченности в поверхностных и подземных водах. От этого во многом зависят точность получаемых результатов и пределы применимости данного метода. Расчет величин подземного стока $Q_{под}$ проводится по общеизвестной формуле смешения

$$Q_{под} = Q_р \frac{C_р - C_{пов}}{C_{под} - C_{пов}}, \qquad (2.1)$$

где Q_p – расход реки на расчетную дату; C_p – концентрация гидрокарбонат-иона в речных водах (используются фактические значения либо снимаются с кривой зависимости этого ингредиента от расходов рек); $C_{пов}$ – концентрация гидрокарбонат-иона в поверхностных водах; $C_{под}$ – содержание гидрокарбонат-иона в подземных водах. Две последние характеристики обосновываются их вероятностно-статистической обработкой с использованием кривых продолжительности и обеспеченности концентраций маркирующего ингредиента.

При подготовке исходных данных для моделирования значительные изменения претерпели лишь величины коэффициентов подземного стока, что связано с исправлением и экстраполяцией данных по увлажнению в речных бассейнах и высотных зонах.

2.2. ГЕОЛОГО-СТРУКТУРНЫЕ ПРИЗНАКИ

В качестве основных признаков, характеризующих геолого-структурные особенности территории, приняты плотность тектонических нарушений кайнозойской активизации Байкальской рифтовой зоны (α) и относительный показатель скорости вертикальных движений земной коры. Выбор признаков обусловлен тем, что в кайнозое (неогеновый период) произошли активные тектонические движения, проявлявшиеся преимущественно в разрывных нарушениях. Движения привели к формированию современного рельефа и, следовательно, существующей геологической обстановки.

Разрывные смещения наряду с физико-механическими свойствами горных пород оказывают большое влияние на гидрогеологические условия территории. В Прибайкалье отчетливо выделяются два главных направления разрывных смещений – северо-восточное и северо-западное. Эти разломы – главные структурные линии региона и протягиваются от оз. Хубсугул к северной оконечности Верхне-Ангарской впадины (Обручевский сброс). Генеральная ветвь состоит из серии разломов различного генезиса, а все другие смещения так или иначе связаны с Обручевским сбросом /Пиннекер и др., 1968/.

Обширный фактический материал по Монголо-Байкальскому орогену свидетельствует о том, что в горно-складчатых системах преобладают плотные кристаллические породы, литологический состав которых определяет степень и характер трещиноватости и раздробленности горных пород, формирующихся и обновляющихся в результате активных тектонических движений. Поэтому в пределах речных бассейнов условия формирования подземного стока во многом определяются такими характеристиками, как плотность тектонических нарушений /Естественные ресурсы..., 1976/. Под последней понимается количество разломов земной коры, отнесенное к единице площади. Методика расчета этого показателя подробно изложена в работе С.И. Шермана /1977/.

Относительный показатель, выраженный через градиент скорости вертикальных движений земной коры, представляет изменение тангенса угла наклона поверхности некоторого участка земной коры в единицу времени. Его значение можно определить по следующей формуле /Гзовский, 1971, 1975/:

$$/\mathrm{grad}\ V/_m = \frac{\Delta h_1 - \Delta h_2}{lt}\ \text{год}^{-1}, \qquad (2.2)$$

где Δh_1 и Δh_2 – соответственно максимальное и минимальное значения поверхности, снятые с карт осредненного рельефа; l – диаметр круговой па-

летки (50 км), с помощью которой производилось построение карты, выбор размера окна диктовался мощностью земной коры в регионе; t – расчетный период времени (35 млн. лет для Сибирской платформы и Байкальской рифтовой зоны и 10 млн. лет для Забайкальской плиты). Прямые скобки указывают, что рассчитываемый модуль – абсолютная величина, а индекс m обозначает среднее значение величины градиента.

Данные о плотности тектонических нарушений кайнозойской активизации и величины относительного показателя скорости вертикальных движений земной коры для речных бассейнов определялись с одноименных карт, построенных К.Г. Леви /1981/. В основу построения карты скоростей неотектонических движений положено представление о рельефе земной поверхности как о едином экзогенно-активном слое. Кровлей этого слоя является обобщенная вершинная, а подошвой – базисная поверхность. Выявляются поверхности нулевого порядка (Сибирская платформа и Забайкальская плита) и производится последовательное сглаживание рельефа, что позволяет снять экзогенную составляющую и выявить кайнозойские тектонические структуры различных рангов. Относительный показатель скорости вертикальных движений земной коры характеризует наклон основных геоструктурных элементов, а плотность тектонических нарушений – их естественную раздробленность.

По мнению М.В. Гзовского /1971/, карты средней величины градиента скорости вертикальных движений земной коры – вполне объективные документы, они наиболее надежны для областей новейшей активной тектоники (Байкальская рифтовая зона – БРЗ). В работе К.Г. Леви /1981/ отмечается тесная связь показателя градиента скорости с сейсмичностью региона и, следовательно, с разрывной тектоникой, создающей потенциальные возможности для проникновения и циркуляции подземных вод в земной коре. Достоинство градиента вертикальных неотектонических движений земной коры состоит в том, что им дополнительно учитывается роль ряда геолого-структурных особенностей региона. В совокупности указанные признаки для горно-складчатых областей отражают региональные особенности коллекторских свойств водовмещающих пород и оказывают непосредственное влияние на пространственное распределение ресурсов подземных вод.

2.3. МОРФОМЕТРИЧЕСКИЕ И ГЕОМОРФОЛОГИЧЕСКИЕ ПРИЗНАКИ

В работах В.А. Всеволожского /1973, 1983/, И.Н. Гарцмана и др. /1973/, И.С. Зекцера /1977/, В.С. Ковалевского /1976/, Б.И. Куделина /1960/, О.В. Попова /1968/, Н.А. Ржаницына /1960/, Б.Л. Соколова, В.О. Саркисяна /1980/, Р.Е. Хортона /1948/, Б.М. Шенькмана /1978/ и других убедительно показано влияние морфометрической и геоморфологической обстановки на распределение и формирование подземных водных ресурсов в различных регионах.

Для построения многомерных моделей формирования подземного стока нами использованы следующие признаки: площадь речного водосбора (F); средняя отметка речного русла ($H_р$); средняя высота речного бассейна ($H_б$); средний уклон бассейна ($I_б$); средний уклон русла ($I_р$); расстояние от замыкающего створа до наиболее удаленной точки бассейна (L). Некоторые из этих признаков выбраны из соответствующих справочников по водным ресурсам, а частью рассчитаны. Средние отметки и уклоны русел рассчитывались методом средневзвешенного по их показателям, затем определялись для замыкающих гидрометрических створов в пределах высотных поясов. Ошибка определения признаков составила 10-12%.

В качестве показателей горизонтального и вертикального расчленения рельефа приняты: частота потоков (S), представляющая отношение количества всех потоков различных порядков к площади водосбора; густота речной сети (D), рассчитываемая как отношение суммарной длины потоков к площади речного водосбора; мощность эрозионного вреза рек ($h_{вр}$), определяемая как разность между средней высотой водосбора и средней отметкой речного русла; комплексный показатель расчленения рельефа (K). Последняя характеристика представляет собой произведение удвоенной густоты речной сети на мощность эрозионного вреза рек и отражает в совокупности как горизонтальное, так и вертикальное расчленение рельефа речного бассейна.

Кроме перечисленных признаков, для расчетов использованы такие морфометрические признаки, как суммарная длина гидросети в пределах речного бассейна (l_o), порядок магистрального водостока, на котором расположен гидрометрический створ (N), и длина магистрального водотока (l_n). Расчет густоты речной сети, порядка и частоты потоков осуществлялся по схемам гидрографической сети и данным, опубликованным в работе /Гидрологическая изученность, 1966/. Оценка порядков водотоков произведена по классификации Р.Е. Хортона /1948/. Согласно этой классификации, самые малые, неразветвленные водотоки относятся к первому порядку, следующие, принимающие в себя потоки первого порядка, - ко второму и т.д., вплоть до главной реки, относимой к самому высшему порядку, характеризующему одновременно порядок всей речной системы. Преимущество этой классификации перед другими подобными заключается в простоте и нетрудоемкости расчетных схем.

Отклонение вычисленных значений густоты речной сети от величин аналогичных параметров, опубликованных в материалах Госкомгидромета СССР, не превышает 20-25%.

2.4. ГИДРОМЕТЕОРОЛОГИЧЕСКИЕ ПРИЗНАКИ

Атмосферные осадки и испарение - наиболее важные метеорологические характеристики, оказывающие влияние на процессы распределения подземного стока. Нет необходимости останавливаться на значении правильности и объективности их расчетов при моделировании природных процессов. Корректность расчетов этих признаков, а также величин естественных ресурсов подземных вод определяет характер основных статистических связей в системе атмосферные осадки - подземный сток. Для Байкальского региона существует довольно много карт распределения атмосферных осадков. Общий недостаток этих схем - отсутствие согласованности с другими элементами водного баланса (речным стоком и испарением). Такая попытка сделана в работе А.Н. Афанасьева /1976/, но величины испарения в этой монографии рассчитаны по разнице уравнения водного баланса. Наиболее достоверно (с нашей точки зрения) для подавляющего большинства метеостанций в последние годы проведен расчет гидроклиматических показателей В.С. Мезенцевым и Г.С. Ратушняком. Все эти исследователи при построении карт увлажнения Байкальского региона столкнулись с весьма значительными трудностями, обусловленными слабой гидрометеорологической изученностью, особенно в высокогорной зоне.

Для получения расчетных данных элементов водного баланса отдельных речных бассейнов использовались величины общего увлажнения по 123 метеостанциям за многолетний период. На основании пространственного корреляционного анализа между рядами годовых сумм атмосферных осадков вся территория разделена на восемь районов, характеризующихся общими закономерно-

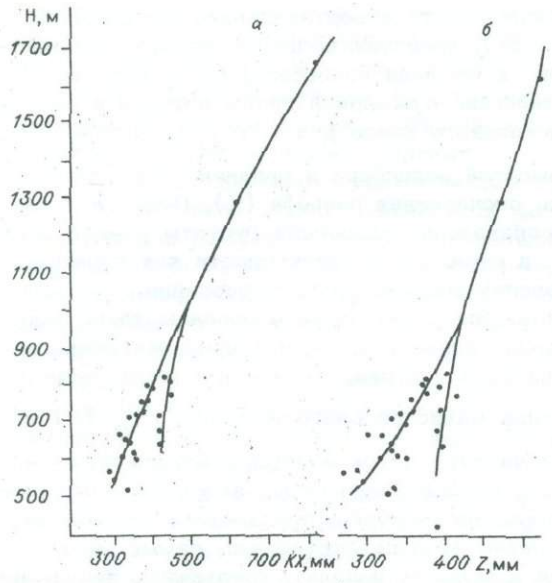


Рис. 4. Зависимости среднемноголетних величин общего увлажнения (а) и испарения (б) от высот местности для центральной части Селенгинской Даурии.

стями распределения увлажнения. Методика пространственного корреляционного анализа метеорологических данных подробно изложена в работе Л.С. Гандина и Р.С. Кагана /1976/.

Территориально выделенные районы расположились следующим образом (в скобках приводится диапазон изменения величин общего увлажнения в миллиметрах):

1 – северо-западный склон хр. Хамар-Дабан (600–1500); 2 – западный склон Баргузинского хребта и северо-западный склон хр. Улан-Бургасы (350–600); 3 – юго-восточный склон Приморского хребта (250–400); 4 – восточный склон Байкальского хребта севернее р. Рель и Верхнеангарская впадина (400–1300); 5 – Баргузинская впадина и ее обрамление (250–600); 6 – Юго-Западное Забайкалье – бассейны рек Джида и Темник (300–600); 7 – верховья бассейнов рек Уда и Хилок (250–500); 8 – основная часть Селенгинской Даурии, включая бассейн р. Чикой (250–500).

Выделенные районы при дальнейшем дроблении можно разбить на более мелкие таксономические подразделения. Предлагаемая схема распределения общего увлажнения в целом согласуется с районированием по распределению элементов водного баланса в зависимости от влияния форм рельефа и направления влагоносных ветров /Афанасьев, 1976/.

На следующем этапе работы по каждому из районов строились кривые зависимости величины общего увлажнения и испарения от абсолютных высот метеостанций (рис. 4).

Величина увлажнения включает среднемноголетнюю величину атмосферных осадков, конденсацию водяных паров воздуха, не учитываемую осадкомерными приборами, росу, иней, изморозь.

С использованием зависимостей через среднюю высоту бассейнов рек определены среднемноголетние величины общего увлажнения и испарения для речных бассейнов, по которым составлялось элементарное уравнение водного баланса

$$KX = Y - Z = \pm \Delta, \qquad (2.3)$$

где KX – общее увлажнение, Y – речной сток, Z – испарение, $\pm\Delta$ – невязка уравнения водного баланса.

Величина невязок уравнения водного баланса на первом этапе расчетов для отдельных бассейнов достигала ± 200 мм, что объясняется недостаточной надежностью некоторых полученных зависимостей общего увлажнения и испарения от высоты местности; зачастую такие кривые отстраивались по 5–6 точкам. В подавляющем большинстве значения невязок не превышали первых десятков, что свидетельствует о достаточно высокой точности расчетов основных

компонентов водного баланса и правильности выделенных районов. Специальный анализ изменчивости этих компонентов показал, что наибольшая вероятность погрешности связана с пространственной изменчивостью увлажнения в регионе, и именно этот показатель нуждается в дальнейшем уточнении. Величина увлажнения, полученная расчетным путем в бассейне оз. Байкал, изменяется от 210 до 1720 мм, испарение – от 200 до 780 мм; для речного стока эти изменения составляют от 21 до 983 мм.

Средняя величина погрешности расчетов речного стока минимальна по сравнению с другими элементами водного баланса и по данным справочников для гидрометрических постов региона составляет около 6%. Величина общего увлажнения корректировалась на основе уравнения водного баланса В.С. Мезенцева /1973/, которое после преобразования имеет следующий вид:

$$\frac{KX}{Z} = \left[1 + \left(\frac{Z_{max}}{KX + W_2 - W_1}\right)^n\right]^{-\frac{1}{n}} + \frac{Y}{Z_{max}}, \qquad (2.4)$$

где KX – общее увлажнение территории, равное измеренным осадкам X, умноженным на К-коэффициент, учитывающий инструментальные погрешности устаревших конструкций осадкомеров, недоучет таких статей прихода влаги, как конденсация, иней, изморозь, роса и ветровое выдувание; Z_{max} – максимально возможное испарение; W_1, W_2 – запасы влаги в почвенном слое соответственно в начале и конце расчетного периода; Y – речной сток; n – безразмерный параметр ($n = 2$ для горно-складчатых областей, $n = 3$ для равнинных территорий).

Поскольку в многолетнем разрезе $W_2 - W_1 \approx 0$, то (2.4) запишется

$$\frac{KX}{Z} = \left[1 + \left(\frac{Z_{max}}{KX}\right)^n\right]^{-\frac{1}{n}} + \frac{Y}{Z_{max}}, \qquad (2.5)$$

а соотношение

$$\frac{Y}{Z_{max}} = f\left(\frac{KX}{Z_{max}}\right). \qquad (2.6)$$

Иными словами, поверхностный сток представлен функцией от величины общего увлажнения и максимально возможного испарения.

Как и все другие аналитические выражения водного баланса, уравнение (2.5) обладает компенсационными свойствами. Так, по данным Н.С. Беркина /1970/, Н.С. Беркина и В.А. Малия /1970/, для районов оптимального увлажнения Прибайкалья при ошибках до 20% в расчетах величин максимального испарения ошибка определения величины испарения по зависимости вида $\frac{Y}{Z_{max}} = f\left(\frac{KX}{Z_{max}}\right)$ составит 8-10%, а для условий недостаточного увлажнения – 4-6%. С помощью этой зависимости, используя компенсационные связи уравнения водного баланса, можно получить коррективы для величин общего увлажнения.

Окончательные расчеты величин увлажнения для речных бассейнов проводились следующим образом:

а) для всех речных бассейнов рассчитана средневзвешенная величина максимально возможного испарения, и на основании данных по речному стоку определено отношение $\left(\frac{Y}{Z_{max}}\right)$;

б) для выделенных по величине увлажнения районов строились зависимости вида $\frac{Y}{Z_{max}} = f\left(\frac{KX}{Z_{max}}\right)$;

в) по величине $\left(\frac{Y}{Z_{max}}\right)$ определено новое отношение $\left(\frac{KX}{Z_{max}}\right)$ и затем путем умножения на значение (Z_{max}) получена откорректированная величина увлажнения в бассейнах рек.

В результате таких операций величина невязок, полученных по уравнению водного баланса с использованием исправленных величин общего увлажнения, заметно уменьшилась. Максимальные значения невязок лишь в единичных случаях достигали 100 мм. Для большинства речных бассейнов значение этой характеристики не превышает 20-40 мм.

Кроме указанных методических приемов контроль за правильностью определения составляющих водного баланса и расчетов подземного стока осуществлялся с использованием гидрохимической информации /Хаустов, 1982в; Хаустов, Рассохин, 1982; Pisarskii, Khaustov, 1981/. На основе предлагаемых в этих работах методов можно не только выявить роль отдельных составляющих водного баланса в круговороте вещества в речных бассейнах, но и решить обратную задачу - определить отдельные компоненты водного баланса. Проведенные расчеты по основным рекам бассейна оз. Байкал показали правильное соотношение водно-солевого баланса; это позволяет утверждать, что для дальнейшего моделирования используются в целом достаточно надежные данные по увлажнению, испарению и параметрам подземного стока. Дополнительно к перечисленным климатическим признакам в расчетную схему факторного анализа включены такие характеристики, как величина максимально возможного испарения, невязки уравнения водного баланса и коэффициенты естественной зарегулированности речного стока.

Коэффициенты естественной зарегулированности отражают базисный сток реки, который сформирован преимущественно за счет подземного питания при небольшой доле поверхностного стока. Формирование последней вызвано преимущественно такими аккумуляционными факторами, как озерность, заболоченность, наличие площадей развития многолетне- и сезонномерзлых пород и достаточно больших размеров водосборных площадей. Коэффициент естественной зарегулированности представляет собой площадь кривой продолжительности, или площадь гидрографа стока до среднего расхода реки. Аналитически он может быть выражен следующим соотношением:

$$\varphi = \oint_0^1 p\,dk, \qquad (2.7)$$

где p - обеспеченность в долях от единицы; k - модульный коэффициент, или отношение единичного расхода к среднему. Коэффициент естественной зарегулированности - величина безразмерная.

2.5. ПРИЗНАКИ ВЫСОТНОЙ ПОЯСНОСТИ ПОДЗЕМНОГО СТОКА

Исследованиями последних лет для горно-складчатых сооружений выявлена гидрогеодинамическая поясность формирования ресурсов подземных вод /Писарский, Хаустов, Костин, 1975; Писарский, Хаустов, 1979; Хаустов, 1982а/. Для Монголо-Байкальского орогена в пределах речных бассейнов выделены приводораздельный, склоновый и долинный пояса, отличающиеся своеобразием в формировании и пространственно-временном распределении подземного стока. Такая поясность носит региональный характер не только для горных сооружений Прибайкалья, где она была выявлена, но и для других орогенных образований.

В приводораздельном поясе, который охватывает верхнюю часть горных хребтов, подземный сток имеет временный характер, более динамичен в силу больших уклонов местности и отсутствия мощных толщ аллювиальных отложений, характеризуется максимальной изменчивостью. Для одного из районов Юго-Западного Забайкалья в период наибольшей динамичности в летне-осенний сезон реализуется в речную сеть до 70% годового количества подземного стока. В лимитирующий сезон (декабрь - март) подземный сток составляет в приводораздельном поясе около 4% годового количества. На формирование подземного стока в этих районах расходуется до 5% величины увлажнения, такой низкий коэффициент во многом обусловлен наличием в Байкальском регионе многолетнемерзлых пород, которые по мере увеличения абсолютных отметок местности имеют большее распространение и бóльшую мощность. Последняя играет роль водоупора, и бóльшая часть атмосферных осадков реализуется поверхностным стоком /Хаустов, 1979а; Хаустов, 1981/.

В склоновом поясе подземный сток существует в течение всего периода, но в зимнее время значительно истощается и своим существованием обязан ресурсам подземных вод, формирующимся преимущественно в зонах крупных тектонических нарушений.

В условиях долинного пояса, несмотря на увеличение испарения в водном балансе, подземный сток более устойчив. Этому способствует широкое развитие чехла осадочных рыхлых образований с большой мощностью и аллювиальных хорошо проницаемых отложений. Мощная зона дренирования и незначительные уклоны местности обеспечивают равномерное поступление подземных вод в речную сеть. Коэффициенты вариации подземного стока долинного пояса в 3-4 раза ниже таковых для условий приводораздельного пояса.

Выделение поясов подземного стока осуществлялось на топографических картах различного масштаба - от 1:100 000 и мельче. Масштаб выбирался в зависимости от площади речных бассейнов: средним водотокам с площадью бассейна от 100 до 10 000 км^2 соответствовал 1:100 000 и мельче. Таким образом на первом этапе выделения поясов отстраивались продольные профили русел рек, а затем на основе анализа геолого-гидрогеологической обстановки намечались их границы. В каждом отдельном случае фиксировались данные о рельефе местности, форме речных бассейнов, уклонах, порядках, длинах и частотах водотоков, морфометрии русел рек, глубинах эрозионных врезов водотоков, наличии наледей в зимний период и других косвенных признаках. Их учет базировался на результатах полевых и расчетных исследований в репрезентативных речных бассейнах различных ландшафтных зон в пределах бассейна оз. Байкал как в СССР, так и в МНР.

На втором этапе работы на топооснове по горизонталям проводились границы поясов и осуществлялся расчет следующих морфометрических и геоморфологических показателей для каждого пояса: площадей ($F_{вод}$, $F_{скл}$,

$F_{дол}$); средних уклонов водосборов ($I_{б\ вод}$, $I_{б\ скл}$, $I_{б\ дол}$); средних высот поясов ($H_{вод}$ $H_{скл}$, $H_{дол}$); коэффициентов густоты речной сети ($D_{вод}$, $D_{скл}$, $D_{дол}$); уклонов речного русла ($I_{р\ вод}$, $I_{р\ скл}$, $I_{р\ дол}$). Приближенная оценка точности перечисленных признаков показала, что ошибка их измерения в редких случаях превышает 10%.

Таким образом, в основу построения факторной модели на первом этапе было положено всего 37 исходных признаков, как прямо, так и косвенно характеризующих основные закономерности формирования естественных ресурсов подземных вод в Байкальской горно-складчатой системе.

Глава 3

ФАКТОРНЫЕ МОДЕЛИ ФОРМИРОВАНИЯ ЕСТЕСТВЕННЫХ РЕСУРСОВ ПОДЗЕМНЫХ ВОД ГОРНО-СКЛАДЧАТЫХ ОБЛАСТЕЙ

3.1. МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ И ОСНОВЫ ФАКТОРНОГО АНАЛИЗА

В предыдущих главах отмечалось, что при анализе природных систем, в частности гидрогеологических, приходится иметь дело с диффузным плохо организованным характером протекания процессов формирования подземных вод в этих системах. Случайный характер процессов позволяет дифференцировать гидрогеологические системы посредством анализа статистических свойств их взаимосвязей. В силу влияния на эти процессы комплекса природных условий исследователь сталкивается с большим количеством исходной информации, которую трудно оценить без привлечения современных методов многомерного статистического анализа. К разряду таких методов относятся: факторный (метод главных компонент), кластерный, регрессионный анализы и др.

Применительно к решению геологических задач методы многомерного анализа использовались сравнительно давно /Абрамович и др., 1975; Белонин и др., 1982; Бугаец, Дуденко, 1976; Воронин, 1982; Девис, 1977; Иёреског и др., 1980; Крамбейн, Грейбилл, 1969; Крамбейн и др., 1973; Миллер, Кан, 1965; Temple, 1978; и др./. Дополнительные сведения по теоретическим основам методов многомерного (и, в частности, факторного) анализа детально изложены в ряде работ /Айвазян и др., 1974; Андерсон, 1963; Лоули, Максвелл, 1967; Окунь, 1974; Харман, 1972; Jahn, Vahle, 1970; Weber, 1974; и др./, в некоторых приводятся алгоритмы, с помощью которых возможна реализация методов на ЭВМ, поэтому остановимся на самых общих свойствах мощного аппарата исследования и проблемах, возникающих на практической стадии его использования.

В гидрогеологических исследованиях вероятностно-статистические методы довольно прочно завоевали позиции, и особенно эффективно их приложение к задачам прогноза временной изменчивости количественных показателей ресурсам подземных вод. В обобщающих работах А.А. Коноплянцева и С.М. Семенова /1974/, В.С. Ковалевского /1976, 1983/, Э.А. Зальцберга /1976/, Б.Л. Соколова и В.О. Саркисяна /1980/, Б.А. Шмагина /1981/ и других, наряду с традиционным использованием вероятностно-статистического аппара-

та корреляционного и регрессионного анализов, широко применяются методы выборочных распределений и множественной регрессии. Методы факторного анализа в концепциях многомерного статистического анализа начали применяться сравнительно недавно в гидрогеологических исследованиях, и число работ, посвященных этим методам, пока еще ограниченно и не всегда приводит к удачному решению поставленных задач.

В одной из последних работ В.С. Ковалевского /1983/ показано, что применение факторного анализа для районирования подземного стока по особенностям его многолетнего режима на территории СССР не дало положительных результатов, хотя здесь же приведен частный пример, где применение метода весьма эффективно. В качестве положительных примеров использования методов факторного анализа (главных компонент) при различного вида гидрогеолого-гидрогеохимических исследованиях можно сослаться на ряд работ /Белонин и др., 1982; Григорьева, Чесалов, 1981; Йёреског и др., 1980; Коротков, 1983; Лисенков, Денисов, 1977; Павличенко, Шапиро, 1982; Писарский, Хаустов, 1982; Хаустов, 1978; Хаустов, Федоров, 1979; Хаустов, Чесалов, 1981; *Хаустов, 1984а; Шапиро и др., 1982; и др.*/.

Сложные комбинации признаков, оказывающих влияние на пространственное распределение ресурсов подземных вод в горно-складчатых областях, не всегда удается выделить в чистом виде и изучить изолированно. Наилучший путь в данном случае - сокращение признакового пространства, т.е. выделение *группы наиболее информативных признаков, несущих информацию о процессах* формирования ресурсов подземных вод. Одно из основных условий при подобном подходе - незначительные ее потери. Наиболее эффективный способ решения этих задач - метод факторного анализа или его разновидность - метод главных компонент. Применение методов в современной литературе основано *на линейном преобразовании* по ортогональным осям в многомерном пространстве.

Основную гипотезу факторного анализа Дж. Девис /1977/ формулирует следующим образом. В природе данные наблюдений лишь косвенным образом характеризуют изучаемое явление, на самом деле существуют скрытые ненаблюдаемые параметры или свойства, число которых ограниченно. Они и определяют основные свойства статистической совокупности рядов наблюдений. Отсюда вытекает одно из основных достоинств факторного анализа - он свободен от каких-либо априорных предположений об ожидаемых результатах.

В основу построения факторных моделей формирования ресурсов подземных вод положены следующие предпосылки /Хаустов, 1979б; 1984а, б/.

1. Процесс формирования подземного стока представляет совокупность различных состояний природной обстановки, он определяется сравнительно небольшим числом характеристик (признаков).

2. Признаки, оказывающие наибольшее влияние на формирование подземного стока, так же изменчивы во времени, как и подземный сток. Возможны лишь некоторые временные сдвиги, обусловленные инерционностью процессов его формирования.

3. Под фактором понимается совокупность присущих только ему взаимосвязанных признаков, оказывающих влияние на пространственно-временную структуру подземного стока.

4. Число признаков должно быть значительно больше количества факторов, признаки внутри факторов распадаются на связанные с факторами группы.

5. Роль отдельных признаков в формировании факторов должна быть явно неодинакова, иначе теряется смысл применения методов факторного анализа.

По А.Г. Аркадьеву и Э.М. Браверману /1971/, факторы не представляют собой что-то принципиально отличное от признаков, но они (факторы) не

могут подвергаться непосредственному измерению, и о них можно судить на основании измеряемых величин (характеристик) явлений. Такое определение, с нашей точки зрения, наилучшим образом соответствует предпосылкам, изложенным выше.

В основе современных представлений алгоритмов построения факторных моделей в качестве исходного момента обычно используется метод разложения информации по ортогональным составляющим или метод главных компонент. Их идея /Белонин и др., 1982; Иберла, 1980; Лоули, Максвелл, 1967; и др./ состоит в отыскании некоррелируемых линейных комбинаций вида:

$$Y_j = \sum_{j=1}^{r} A_{i,j} X_i; \; j = 1,2,\ldots, n; \; i = 1, 2,\ldots, m; \; r < m, \qquad (3.1)$$

где X_i – исходная информация, заданная в точках поля или в виде матрицы исходных данных (признаков с числом m); $A_{i,j}$ – некоторые коэффициенты (факторные нагрузки). Необходимо найти такие значения коэффициентов $A_{i,j}$, при которых дисперсии Y_j достигают экстремальных значений. Уравнение (3.1) постулирует одно из основных предположений факторного анализа о том, что множество коррелируемых рядов наблюдений (X_i) можно описать меньшим числом комплексных переменных – факторов ($A_{i,j}$). В геометрическом представлении эта задача интерпретируется как переход к новой системе координатных осей, на которые падают максимальные дисперсии признаков.

Вычислительные операции сводятся к решению уравнения вида

$$(R - \lambda E)A = O, \qquad (3.2)$$

где R – ковариационная, а при условии стандартизации исходных данных – *корреляционная матрица*; E – *единичная матрица*; A – *собственные векторы* (главные компоненты, факторы); λ – собственные значения (числа) матрицы R, отвечающие дисперсии линейных преобразований Y_j. Это уравнение означает, что неизвестный вектор A ортогонален всем векторам строки матрицы $(R-\lambda E)$. Иными словами, решение основного уравнения компонентного анализа сводится к нахождению собственных векторов и значений ковариационной (или корреляционной) матрицы. Максимальные дисперсии признаков выражаются через собственные значения (λ) корреляционной матрицы (R) и являются показателями разброса признаков вдоль главных координантных осей. Собственные векторы (A) определяют направление главных осей, соответствующих максимальной дисперсии признаков. Задача нахождения собственных векторов рассматривается как процедура вращения координатных осей с целью максимизации дисперсий признаков. В случае наличия связи между двумя или большим числом признаков возрастает удельный вес собственных векторов. Это замечательное свойство дает возможность представить метод главных компонент как эффективный аппарат свертки исходной информации при переходе к новой ортогональной системе координат. Собственные векторы представляют косинусы углов поворота осей от старой системы координат к новой и являются независимыми линейными комбинациями исходных признаков.

Величина собственных значений (λ) корреляционной матрицы убывает по направлению от главной (первой) компоненты к последней, а их сумма равна сумме диагональных элементов корреляционной матрицы, следу, или ее порядку, который определяется по числу признаков. Вклад компоненты в об-

щую дисперсию признаков обычно выражается в процентах и определяется с помощью соотношения

$$\frac{\lambda_i}{\sum\lambda} = \frac{\lambda_i}{tr(R)}, \qquad (3.3)$$

где $tr(R)$ - след (порядок) корреляционной матрицы.

Под факторной нагрузкой понимается величина, аналогичная коэффициенту корреляции между признаками и главными компонентами, которая представляется в виде матрицы факторных нагрузок (W). Сумма квадратов элементов строки этой матрицы характеризует дисперсию признаков, описанную компонентами или факторами

$$\sum_{j=r}^{r} A_{i,j}^2 = 1, \text{ где } j = 1, 2, \ldots, n. \qquad (3.4)$$

Для компонентного анализа сумма должна быть равной 1, в факторном это требование не обязательно.

Суммы квадратов факторных нагрузок матрицы (W) по столбцам равны собственному значению исходной матрицы (R) и являются дисперсией главных компонент

$$\sum_{i=1}^{r} A_{i,j}^2 = \lambda \text{ и } \sum_j \sum_i A_{i,j}^2 = \sum_j \lambda \qquad (3.5)$$

Приведенные выражения показывают, что квадрат факторной нагрузки компоненты на i-й признак отражает вклад компоненты в описание i-го признака.

В силу принципа ортогональности выделенных компонент выявляется еще одно важное свойство матрицы факторных нагрузок - сумма попарных произведений факторных нагрузок, соответствующая двум факторам, равна нулю, а суммирование попарных произведений факторных нагрузок двух признаков на полный набор одноименных факторов есть не что иное, как величина коэффициента корреляции между этими признаками

$$R_{i,l} = \sum_{j=1}^{n} A_{i,j} \cdot A_{lj}. \qquad (3.6)$$

Факторные нагрузки изменяются от -1,0 до 1,0, и так же, как для коэффициентов корреляции при оценке тесноты связей, необходимо проводить оценку их значимости. На основании перечисленных свойств матрицы факторных нагрузок производится ее интерпретация с целью определения компонентной (факторной) структуры и, наоборот - оценка классификации и вклада факторов в зависимости от их признаковой структуры. Последняя задача весьма ответственна, она может быть решена неоднозначно, а интерпретация результатов во многом зависит от вида используемого алгоритма, информации, опыта и эрудиции исследователя.

В основном трудности возникают при оценке необходимого количества компонент, для определения "достаточного" значения суммарной дисперсии в предположении, что число факторов (k) должно быть меньше размерности исходного пространства (m). Д. Лоули и А. Максвелл /1967/ рекомендуют до проведения расчетов использовать соотношение вида

$$E_{(k)} = \frac{(2m+1) \pm \sqrt{(2m+1)^2 - 4m(m-1)}}{2} \qquad (3.7)$$

при условии, что $E(k) \leqslant m$. К. Иберла /1980/, Е. Вебер /Weber,1974/ и Дж. Девис /1977/ предлагают принимать во внимание столько факторов, сколько собственных чисел у корреляционной матрицы больше 1, при этом суммарный вклад числа факторов должен достигать 90%. К.Г. Йёреског и др. /1980/ считают достаточным для описания факторной структуры такое число компонент, при котором их суммарный вклад достигает 95-99%. При этом ставится условие, что величины первых собственных значений корреляционной матрицы (λ) существенно отличаются от нуля. Д.В. Белонин и др. /1982/ при оценке необходимого числа факторов рекомендует брать их не больше собственных чисел со значениями, удовлетворяющими следующему неравенству:

$$\lambda > \frac{\sum \lambda_i}{m}, \qquad (3.8)$$

где m - размерность исходного пространства.

В случае работы с тесно коррелируемыми признаками максимизация факторных нагрузок приходится на первые 3-4 фактора. Этот факт аргументированно продемонстрирован в работе Е. Вебер /Weber, 1974/, именно первые факторы отвечают собственным значениям корреляционной матрицы λ_i больше 1. В этой же работе рекомендуется для оценки числа необходимых факторов строить зависимости собственных значений (λ_i) от номеров факторов. За диагностируемое число факторов, необходимых для интерпретации, принимается точка, в которой кривая выполаживается (имеет резкий перелом), и в дальнейшем значения λ_i изменяются очень мало.

В процессе вычисления матрицы факторных нагрузок можно заранее проверить гипотезу о необходимом (максимальном) числе факторов для описания признакового пространства. Для ограничения их числа (k) в зависимости от числа признаков (m) используется следующее соотношение:

$$(m + k) < (m - k)^2. \qquad (3.9)$$

Определенные трудности интерпретации матрицы факторных нагрузок возникают при использовании разномасштабных единиц измерения признаков. Недоучет этого момента, а также точности определения признакового пространства могут явиться причиной некорректности выводов при интерпретации факторной матрицы. При исследовании естественных процессов чаще всего приходится иметь дело с признаками различной физической природы, поэтому для преодоления этой трудности их стандартизируют.

Эта процедура сводится к вычитанию среднего арифметического из каждого значения признаков с последующим делением на стандартное отклонение, в этом случае данные показывают, на сколько стандартных отклонений каждое случайное наблюдение отличается от среднего, и позволяют довольно легко (по экстремальным значениям) выявить ошибки измерений того или иного признака. Кроме этого, стандартизированные данные, используемые для построения моделей данных, можно применять при оценке соответствия нормальному закону распределения. В случае асимметрических выборок, которые возникают вследствие большого количества экстремальных значений признака, можно заменить выборочные значения соответствующими им логарифмами. Для отрицательных значений логарифмов ко всем наблюдениям необходимо добавить ту или другую константу. Различные способы преобразования информации можно найти в работах Т. Лаудона /1981/, Р. Миллера, Дж. Кана /1965/ и др. Преобразования также полезны в случае нелинейных зависимостей между признаками. Необходимо отметить, что перед применением любого аппарата статистических исследований желательно проводить проверку исходных данных на

нормальность распределения и в зависимости от формы кривой распределения проводить то или иное преобразование.

Как правило, стандартизация данных включает их нормирование, операция нормирования состоит в таком изменении значений векторов наблюдений, чтобы сумма квадратов его элементов была равна 1. Этой процедуре могут быть подвержены как столбцы, так и строки матрицы многомерных данных.

Некоторые исследователи /Йёреског и др., 1980/ считают, что нормирование каждого признака геометрически означает "растягивание" или "сжатие" векторов столбцов матрицы данных, приводящее к единичной длине, что не изменяет их соотношение при изменении пропорциональности в строках. В случае нормирования данных по строкам остаются неизменными пропорции переменных в пределах объекта, но искажаются соотношения.

В основу метода главных компонент положена процедура вращения факторных осей в многомерном пространстве с максимизацией дисперсии признаков на первый и второй ортогональные факторы, затем вновь максимизируется оставшаяся дисперсия. Такие итерации повторяются до тех пор, пока не будет исчерпана основная часть общей дисперсии признаков. Кажущаяся единственность процедуры на самом деле является одним из множеств вариантно-априорных соотношений, достаточно полно описывающих дисперсию признаков в многомерном пространстве. Вполне естественно, что основная цель метода направлена на отыскание минимального геометрического пространства при условии максимизации в нем одновременно всех признаков. Определенная часть информации при этом теряется, особенно в случае большой размерности признакового пространства.

При данных обстоятельствах матрицу факторных нагрузок довольно трудно интерпретировать, так как на первых факторах может быть сосредоточено большое количество зависимых между собой признаков. Такие ситуации наиболее типичны для природных систем, поскольку в них трудно найти полностью "чистые" независимые процессы. Измеряемые нами характеристики, как правило, порождены совместным влиянием целого комплекса процессов иной физической природы. В результате приходится иметь дело с признаками, связанными между собой зависимостями высокого порядка, поэтому на стадии интерпретации результатов расчета факторным анализом возникает неопределенность; каким признакам отдать предпочтение. Здесь возможны два пути к преодолению этой проблемы.

Первый заключается в проведении ряда последовательных просчетов с включением неинформативных и сильно скоррелированных признаков. При исключении хотя бы одного из признаков, вносящих определенный "шум" в общую дисперсию, положение факторных осей изменится. Такими процедурами можно достаточно эффективно увеличить факторные нагрузки других признаков и при этом увеличить долю вклада первых факторов, что имеет большое значение при типологизации объектов в многомерном пространстве.

Второй путь может быть как самостоятельным, так и дополнять первый и заключается в процедуре вращения факторных осей. Основная цель вращения состоит в выборе оптимального положения осей таким образом, чтобы каждая из них содержала небольшое число признаков с максимальными факторными нагрузками. Такие итерации могут быть проведены с помощью специальных критериев, а также относительно каждого признака. В результате вращения получаются факторы, содержащие только комбинации связанных между собой признаков, что значительно облегчает интерпретацию факторной матрицы.

Существуют различные варианты вращения векторных осей, их характеристика детально изложена в работах по факторному анализу.

Дж. Темплом / Temple, 1978/ дан анализ достаточно большого количества работ, в которых используются различные методы вращения осей в

факторном анализе. На основании зачастую умозрительных выводов, не подкрепленных практическими расчетами, этот автор приходит к выводам о субъективности большинства процедур вращения и вообще ставит под сомнение практическое использование факторного анализа в геологии. Следует сразу же отметить, что цитируемые им работы базируются на довольно старых несовершенных алгоритмах, к тому же без применения в комплексе с факторным анализом других статистических методов, позволяющих во многом избежать субъективизма.

Одним из таких методов, позволяющих устранить некоторый произвол при вращении факторных осей, может быть использование искусственных эталонов с заданной статистической структурой /Гавришин, 1980а, б/. Применение таких эталон-моделей позволяет отыскать оптимальное положение факторных осей при вращении, а также выработать достаточно надежные критерии их остановки.

После формирования матрицы факторных нагрузок современными алгоритмами предусмотрен расчет значений главных компонент (факторов) по каждому объекту. Вновь получаемые характеристики для объектов отражают комбинации исходных признаков и могут быть обработаны различными видами статистического анализа. Для расчета этих значений нормированную исходную информацию признаков для каждого объекта умножают на факторные нагрузки, затем суммируют и делят на собственные числа, отвечающие главным компонентам /Белонин и др., 1982/:

$$X_{i,\ n} = \frac{\sum_{i=1}^{k} A_{i,\ j}\, \overline{Y}_{i,\ n}}{\lambda_j}, \qquad (3.10)$$

где $X_{i,\ n}$ значение i-го фактора у n-го наблюдения, $\overline{Y}_{i,\ n}$ – нормированные значения j-го признака у n-го наблюдения, k – число факторов.

Получаемые в результате расчетов по этому выражению данные являются также нормированными величинами, могут быть как отрицательными, так и положительными, их среднее по компоненте (фактору) равно нулю, а дисперсия – 1. Использование таких данных для анализа пространственно-временной изменчивости природных процессов может быть весьма перспективно. Вместе с тем получаемые значения компонент для объектов могут быть использованы для целей классификации, а также для выделения аномальных точек исходной совокупности объектов.

Наиболее наглядно результаты компонентного анализа можно отразить с помощью построения и сравнения карт по отдельным компонентам. Такие карты отражают не только индивидуальные черты пространственной изменчивости отдельных признаков, но и их комбинацию в совокупности. Примеры построения этих карт можно найти в литературе /Хаустов, 1978; Павличенко, Шапиро, 1982; Шапиро и др., 1982/.

До настоящего момента в работе термин "факторный анализ" использовался как синоним термина "главные компоненты". Несмотря на общую основу расчетов, между этими методами имеются существенные различия, основные из которых приводятся ниже. В противоположность компонентному анализу, основанному на максимизации вклада дисперсий, факторный объясняет матрицу ковариацией или корреляцией заранее задним числом факторов, которые находятся по принципу максимизации связей между признаками. Иными словами, факторный анализ имеет дело преимущественно с корреляционными связями, а компонентный – с дисперсиями.

Основная модель факторного анализа может быть записана в следующем виде:

$$Y_j = \sum_{j=1}^{r} A_{i,j} X_j + E_i; \quad i=1, 2,..., m; \; j=1, 2,...,k; (k<m), \quad (3.11)$$

где Y_j – наблюдаемые признаки; m – число признаков; X_j – значение j-го фактора; $A_{i,j}$ – коэффициенты при X_j (нагрузка j-го фактора в i-м признаке); r – количество факторов; E_i – остатки (случайные величины), представляющие собой источники отклонений.

Величины E_i независимы как между собой, так и с факторами и влияют только на наблюдаемые признаки.

Сравнивая аналитические выражения компонентного и факторного анализов, можно видеть их идентичность, отличие же заключается в том, что в последнем уравнении фигурируют остатки. В компонентном анализе предполагается, что этот член уравнения несравненно мал, и относительно его независимости не делается никаких предположений. Последнее предположение о независимости также справедливо в компонентном анализе и относительно исходных признаков, которые могут быть как коррелированы, так и некоррелированы, здесь не требуется условий нормальности их распределения, как в факторном анализе.

Еще одна важная особенность компонентного анализа – его линейность и аддитивность, и относительно этих его свойств не строится никаких гипотез, а при работе с собственно факторным анализом всегда присутствует гипотеза о линейности связей признаков. Эта гипотеза доджна быть всесторонне проверена, и в случае ее неверности, факторную модель можно рассматривать лишь как самое грубое приближение или как один из вариантов множества исходов.

Идентичность решения компонентного и факторного анализов достигается при высоких вкладах первых факторов в общую дисперсию, т.е. когда разница между суммой первых двух и последующих достаточно высока /Белонин и др., 1982; Weber, 1974; и др./. Этими же авторами при исследовании многопараметрических систем в случае отсутствия точных моделей и гипотез отдается предпочтение методу главных компонент, которым придается смысл факторов, и тем самым подчеркивается причинный характер новых гипотетических переменных.

В случае достаточного объема признаков, характеризующих объекты, для их типологизации и анализа взаимосвязанных групп признаков можно применить Q-модификацию факторного анализа. Необходимый объем количества объектов и признаков можно определить, например, с помощью квантилей t-распределения Стьюдента, табличные значения которого приводятся почти во всех руководствах по применению статистических методов. В Q-модификации после нормирования исходных данных транспонируется матрица, т.е. меняются местами столбцы и строки. Иными словами, Q-схема факторного анализа строится на объединении в факторы объектов, которые объединяются в однородные группы с идентичным характером изменчивости набора исходных признаков. Эта модификация (как будет показано ниже) также применима для целей гидрогеологического районирования.

Имеется много различных методов и алгоритмов построения компонентных и факторных моделей. В данной работе для построения факторных моделей использовались два алгоритма, сравнение работы которых на одних и тех же выборках показало их идентичность /Двинских, 1975; Сборник..., 1974/. В основу обоих алгоритмов положен метод главных компонент; достоинство второго алгоритма – возможность осуществления процедуры вращения факторных осей.

В связи с этим можно кратко сформулировать основные теоретические и практические задачи, решаемые с помощью факторного анализа: выявление связей внутри комплекса исходных признаков; выделение групп случайных признаков, обладающих наиболее сильными связями; оценка вклада ведущих признаков и факторов в общую дисперсию; типизация объектов в многомерном пространстве. Последняя задача относится к методам распознавания образов, и здесь очень остро встает вопрос о выборе комплекса наиболее информативных признаков для целей классификации.

3.2. R-ФАКТОРНАЯ МОДЕЛЬ ФОРМИРОВАНИЯ РЕСУРСОВ ПОДЗЕМНЫХ ВОД

Самостоятельное применение факторного анализа как эффективного метода изучения закономерностей формирования ресурсов подземных вод сводится к выявлению и оценке комплекса природных условий, влияющих на пространственное распределение. К числу основных параметров естественных ресурсов подземных вод (по Б.И. Куделину /1960/) относятся модули, коэффициенты подземного стока и подземного питания рек. Распределение этих параметров по площади и во времени происходит под влиянием ряда совместных факторов, которые могут быть идентифицированы через совокупности признаков. Поэтому с системно-кибернетических позиций на первом этапе построения факторной модели необходимо собрать как можно больше исходных данных с тем, чтобы наиболее полно охарактеризовать процесс формирования подземного стока.

Исходя из этих принципов для Байкальской горно-складчатой области, было сформировано признаковое пространство, представляющее матрицу исходных данных размером 80 x 37 (где 80 - число речных бассейнов, 37 - признаки, характеризующие, по нашему мнению, процессы формирования подземного стока в этих бассейнах). Предварительно, на основе гипотетических представлений, исходные данные были объединены в ряд генетически однородных групп (см. гл. 2). Объединение на первом этапе во многом облегчает интерпретацию факторной матрицы. Признаки поясности подземного стока выражались через соотношение их площадей, уклонов, средних высот, густоты речной сети в пределах каждого пояса, к аналогичным показателям для каждого речного бассейна. Такой подход обеспечивает получение равнозначных показателей поясности подземного стока и оценки удельного веса для каждой речной системы. На данном этапе построения факторной модели набор признаков должен был обеспечить объективный подход с точки зрения равноправности причинно-следственных связей.

Расчеты с помощью алгоритма /Двинских, 1975/ позволили выделить 6 главных факторов с суммарным вкладом 67% в общую дисперсию 37 исходных признаков. Эта дисперсия распределяется по факторам следующим образом: на первый - 26%, второй - 13, третий - 9, четвертый - 8, пятый и шестой - 6 и 5% соответственно. Такое относительно равномерное распределение дисперсий признаков на главные факторы свидетельствует либо о разнонаправленности процессов формирования ресурсов подземных вод, либо о том, что исходная информация содержит большое количество шумов, не имеющих отношения к рассматриваемым процессам или слабо влияющих на них.

Анализ матрицы факторных нагрузок (табл. 3) производится с целью интерпретации и выявления взаимосвязанных признаков. Наибольшие нагрузки на первый фактор имеют среднемноголетние расчетные параметры подземного стока: модуль подземного стока ($r = 0,84$), коэффициент подземного стока ($r = 0,85$) и коэффициент подземного питания ($r = 0,69$). Аналогично высокие

Таблица 3. Матрица факторных нагрузок признаков

Признак	Фактор					
	1	2	3	4	5	6
Площадь бассейна (F)	-0,37	0,33	-0,31	-0,36	-0,14	0,00
Средняя высота бассейна ($H_{ср}$)	0,37	0,60	0,50	-0,30	0,09	0,07
Средний уклон русла ($I_р$)	0,73	-0,18	0,17	0,12	-0,06	0,28
Средний уклон бассейна ($I_б$)	0,80	-0,11	0,11	-0,10	0,00	0,04
Расстояние до наиболее удаленной точки бассейна (L)	-0,42	0,38	-0,32	-0,40	-0,16	0,00
Порядок магистрального водотока (N)	-0,60	0,47	-0,17	-0,40	0,10	-0,09
Частота потоков (S)	0,65	-0,27	0,14	-0,43	-0,34	-0,09
Густота речной сети (D)	0,72	-0,22	0,16	-0,34	-0,44	-0,01
Средняя отметка русла ($H_р$)	0,02	0,30	0,79	-0,19	-0,09	-0,05
Мощность эрозионного вреза рек ($h_{вр}$)	0,58	0,58	-0,20	-0,24	0,26	0,25
Комплексный показатель расчленения рельефа (K)	0,80	0,25	-0,01	-0,40	-0,04	0,11
Объем зоны дренирования ($V_{з.д}$)	-0,18	0,17	-0,03	-0,45	-0,15	0,05
Суммарная длина гидросети в бассейне ($l_о$)	-0,67	0,16	-0,11	0,29	0,48	0,03
Длина порядка магистрального водотока ($l_п$)	0,01	-0,15	-0,26	0,57	0,19	0,39
Модуль подземного стока ($M_{п.с}$)	0,84	-0,12	-0,29	-0,02	0,16	-0,06
Коэффициент подземного питания ($K_{п.п}$)	0,69	-0,15	-0,48	0,16	-0,08	0,04
Коэффициент подземного стока ($K_{п.с}$)	0,85	-0,07	-0,36	0,05	0,08	0,02
Общее увлажнение (KX)	0,84	0,09	0,10	-0,14	0,19	-0,16
Суммарное испарение (Z)	0,67	0,00	0,01	-0,32	0,28	-0,39
Максимально возможное испарение (Z_{max})	-0,09	-0,51	0,14	0,17	-0,34	-0,10
Невязка водного баланса (Δ)	-0,15	0,04	0,13	-0,39	-0,30	-0,32
Коэффициент естественной зарегулированности речного стока (φ)	0,43	-0,29	-0,47	0,06	-0,14	-0,10
Залесенность речных водосборов (C)	0,03	-0,60	0,34	-0,02	-0,21	-0,43
Плотность тектонических нарушений (α)	0,52	0,09	-0,48	0,14	0,37	0,07
Относительный градиент скорости вертикальных движений земной коры (γ)	0,75	-0,05	-0,42	0,11	0,14	-0,11
Площадь пояса						
приводораздельного ($F_п/F$)	0,33	0,60	0,11	0,51	-0,26	-0,31
склонового ($F_с/F$)	-0,05	-0,52	0,28	-0,50	0,41	0,05
долинного ($F_д/F$)	-0,40	-0,04	-0,51	0,04	-0,32	-0,06
Высота пояса						
приводораздельного ($H_п/H$)	0,09	0,75	-0,07	0,01	-0,33	0,09
склонового ($H_с/H$)	-0,48	-0,58	-0,22	-0,37	0,33	0,03
долинного ($H_д/H$)	-0,67	-0,17	-0,28	-0,09	-0,19	-0,43
Уклон пояса						
приводораздельного ($I_п/I_б$)	-0,08	0,77	-0,14	-0,05	-0,15	0,16
склонового ($I_с/I_б$)	-0,18	-0,39	-0,25	-0,18	-0,13	0,06
долинного ($I_д/I_б$)	-0,64	-0,16	-0,30	-0,14	-0,04	0,03
Густота речной сети пояса						
приводораздельного ($D_п/D$)	0,15	0,57	-0,07	0,30	-0,30	-0,36
склонового ($D_с/D$)	0,15	0,06	0,45	0,24	0,20	-0,37
долинного ($D_д/D$)	-0,06	-0,15	0,12	0,015	0,36	-0,64
Вклад факторов в общую дисперсию признаков в %	26	13	9	8	6	5

нагрузки в этом факторе имеют общее увлажнение ($r = 0,84$) и испарение ($r = 0,67$), а также геоморфологические признаки: средний уклон русла ($r = 0,73$), средний уклон бассейна ($r = 0,80$), частота речных потоков ($r = 0,65$), густота речной сети ($r = 0,72$), комплексный показатель расчленения рельефа ($r = 0,80$). Кроме того, большие величины факторных нагрузок свойственны таким признакам, как относительный показатель скорости вертикальных неотектонических движений земной коры ($r = 0,75$) и плотность тектонических нарушений ($r = 0,52$).

Достаточно высокие отрицательные факторные нагрузки в первом факторе имеют: порядок магистрального водотока ($r = 0,60$), суммарная длина гидросети в бассейне ($r = 0,67$), коэффициенты отношения высот долинного пояса ($r = 0,67$) и коэффициент уклонов долинного пояса ($r = 0,64$).

С помощью несложных вычислений, на основании матрицы факторных нагрузок, можно оценить дисперсию признаков, извлекаемую факторами, что также в значительной мере облегчает последующую интерпретацию факторов и придает им определенный смысл. Для этого необходимо возвести все величины факторных нагрузок в квадрат. В табл. 4 рассмотрено лишь три первых фактора, поскольку их вес в сумме дает около 50% общей дисперсии (см. табл. 3). Данные таблицы свидетельствуют, что первый фактор формируют, преимущественно, параметры подземного стока, общее увлажнение и испарение в регионе, показатель скорости вертикальных тектонических движений и признаки расчлененности рельефа, а также уклоны речных бассейнов. Эти признаки имеют приблизительно одинаковую пространственную изменчивость в Байкальской горно-складчатой области.

Таким образом, на основе анализа факторной матрицы и расчета процентной доли признаков, формирующих первый фактор, а также коэффициентов корреляции между признаками (корреляционная матрица из-за громоздкости здесь не приводится) можно сделать вывод, что первый фактор в основном формирует характеристики, выражающие условия питания, движения и разгрузки подземных вод, т.е. собственно подземный сток.

Наибольшие величины факторных нагрузок во втором факторе приходятся на признаки, характеризующие поясность подземного стока: высотные коэффициенты приводораздельного пояса ($r = 0,75$), уклона приводораздельного пояса ($r = 0,77$) и др. Следует отметить, что признаки, характеризующие приводораздельный пояс, имеют положительные факторные нагрузки, а признаки склонового и долинного – отрицательные. Доля изменчивости признаков приводораздельного пояса во втором факторе – от 32 до 59%, а признаков склонового – от 15 до 34%. Это говорит о том, что второй фактор в большей степени сформирован из всех признаков поясности признаками приводораздельного пояса. Значительные нагрузки на этот фактор также приходятся на среднюю высоту речных бассейнов ($r = 0,60$), мощность эрозионного вреза рек ($r = 0,58$), залесенность водосборов ($r = 0,60$) и максимально возможное испарение ($r = 0,51$). Вполне естественно, что признаки, характеризующие поясность подземного стока, наряду со средней высотой водосбора и мощностью эрозионного вреза рек, обусловили формирование второго фактора. Их расчеты осуществлялись в основном путем морфоструктурного анализа речных водосборов. В связи с этим второй фактор может быть интерпретирован как фактор, характеризующий комплекс ландшафтных условий и выражений через косвенные признаки поясности подземного стока.

Следует отметить, что один из важнейших показателей дренирования водоносных зон – мощность эрозионного вреза рек – довольно равномерно распределился по двум факторам. В большей степени это обусловлено неточностью расчетов признака. Доказательством является то, что максимальные нагрузки в третьем факторе приходятся на среднюю отметку русла ($r = 0,79$) и сред-

Таблица 4.

Процентные доли признаков, формирующих первые факторы

Признак	Фактор			Сумма по трем факторам
	1	2	3	
1	2	3	4	5
Площадь бассейна (F)	14	11	10	35
Средняя высота бассейна ($H_{ср}$)	14	36	25	75
Средний уклон русла ($I_р$)	53	3	3	59
Средний уклон бассейна ($I_б$)	64	1	1	66
Расстояние до наиболее удаленной точки бассейна (L)	18	14	10	42
Порядок магистрального водотока (N)	36	22	3	61
Частота потоков (S)	42	7	2	51
Густота речной сети (D)	52	5	3	60
Средняя отметка русла ($H_р$)	0	9	62	71
Мощность эрозионного вреза рек ($h_{вр}$)	34	34	4	72
Комплексный показатель расчленения рельефа (K)	64	6	0	70
Модуль подземного стока ($M_{п.с}$)	74	1	8	83
Коэффициент подземного питания ($K_{п.п}$)	48	2	23	73
Коэффициент подземного стока ($K_{п.с}$)	72	0	13	85
Общее увлажнение (KX)	71	1	1	74
Суммарное испарение (Z)	45	0	0	45
Максимально возможное испарение (Z_{max})	1	26	2	29
Невязка водного баланса (Δ)	2	0	2	4
Коэффициент естественной зарегулированности речного стока (φ)	18	8	22	48
Залесенность речных водосборов (C)	0	36	12	48
Объем зоны дренирования ($V_{з.д}$)	3	3	0	6
Суммарная длина гидросети в бассейне (l_o)	45	3	1	49
Длина порядка магистрального водотока (l_n)	0	2	7	9
Плотность тектонических нарушений (α)	27	1	23	61
Относительный градиент скорости вертикальных движений земной коры (γ)	56	0	18	74
Площадь пояса				
приводораздельного ($F_п / F$)	11	36	1	48
склонового ($F_с / F$)	0	27	8	35
долинного ($F_д / F$)	16	0	26	42
Высота пояса				
приводораздельного ($H_п/H$)	1	56	0	57

Окончание табл. 4

1	2	3	4	5
склонового ($H_с/H$)	23	34	5	62
долинного ($H_д/H$)	45	3	8	56
Уклон пояса				
приводораздельного ($I_п/I_б$)	1	59	2	62
склонового ($I_с/I_б$)	3	15	6	24
долинного ($I_д/I_б$)	41	3	9	53
Густота речной сети пояса				
приводораздельного ($D_п/D$)	2	32	0	34
склонового ($D_с/D$)	2	0	20	22
долинного ($D_д/D$)	0	2	1	3

нюю высоту водосбора ($r = 0,50$). Мощность эрозионного вреза рассчитывалась по разности этих двух признаков и, следовательно, должна иметь идентичную дисперсию. Остальные признаки третьего фактора имеют нагрузки менее 0,50, и он может быть интерпретирован как фактор высотного положения речных систем. Отметим, что такая интерпретация носит предварительный характер вследствие перегруженности факторов признаками. Последующие факторы интерпретируются весьма сложно, так как они несут в себе немного информации об исходной совокупности признаков, что отражается малыми величинами вкладов в общую дисперсию.

На данной стадии построения факторной модели мы вплотную столкнулись с проблемой исключения неинформативных признаков, для преодоления которой можно воспользоваться исключением из матрицы исходной совокупности тех признаков, суммарный вклад которых в первые три фактора не превышает 50% (см. табл. 4). К таким признакам относятся: площади речных водосборов и относительные коэффициенты площадей поясов, расстояния до наиболее удаленной точки водосборов, максимально возможное испарение, невязки водного баланса, залесенность речных бассейнов и др.

В этот же набор признаков попадает такая важная характеристика, как испарение с поверхности речных бассейнов. В данном случае необходимо исключение неинформативных признаков рассматривать с позиций их связей с другими признаками, т.е. обращаться в каждом конкретном случае к корреляционной матрице, которую можно наглядно проиллюстрировать с помощью метода Р. Боннера /1969/ или различными видами графов.

Суть метода Боннера заключается в преобразовании матрицы парных коэффициентов корреляции исходной совокупности признаков в матрицу коэффициентов смежности. Последние рассчитываются на основе критических значений для выборочного коэффициента корреляции, которые, в свою очередь, зависят от размеров совокупности коррелируемых рядов и определяют статистическую устойчивость связей. Выборочные коэффициенты корреляции – лишь приближенная оценка истинного коэффициента связи генеральной совокупности, и возникает проблема предельно допустимых отклонений в условиях нулевой или альтернативной гипотезы. Нулевая гипотеза утверждает, что исследуемые случайные величины или события независимы, альтернативная – наоборот.

При расчетах коэффициентов корреляции эти значения в условиях гене-

ральной и выборочной совокупностей могут не совпадать. В этом случае для оценки предельно допустимых отклонений прибегают к проверке соответствия с помощью t-критерия Стьюдента. В условиях нулевой гипотезы распределение Стьюдента с $N-2$ степенями свободы статистика t имеет вид

$$t = \sqrt{\frac{r^2(N-2)}{1-r^2}}, \qquad (3.12)$$

где r – коэффициент корреляции выборочных данных. Из этого выражения легко найти величину допустимого отклонения r_a:

$$/r_a/ = \frac{t_{\alpha, f}}{\sqrt{t^2_{\alpha,f} + (N-2)}}, \qquad (3.13)$$

где $t_{\alpha, f}$ выбирается из таблиц распределения Стьюдента; α – уровень значимости, обычно принимается 0,05 или 0,01; $f = N-2$ – число наблюдений. Как правило, значения $/r_a/$ затабулированы, и их можно найти во многих справочных руководствах по статистике.

Сравнивая значения r и r_a, можно приступить к проверке нулевой гипотезы. Если $r < r_a$, то отклонение носит случайный характер и нет оснований отвергать гипотезу о некоррелированности признаков. В противоположном случае, когда $r \geqslant r_a$, нулевая гипотеза отклоняется при уровне значимости α, а коррелируемые признаки предполагаются линейно связанными. Иными словами, с помощью указанной процедуры можно определить величины значимых коэффициентов корреляции. Для используемой выборки в условиях нулевой гипотезы критическое значение выборочного коэффициента корреляции при 80 членах ряда и уровне значимости $\alpha = 0{,}01$ составляет 0,28.

Воспользовавшись критическими значениями коэффициентов корреляции, довольно легко можно получить матрицу смежности, при построении которой всем величинам коэффициентов корреляции ниже этой приписывается 0, а превышающим – 1. Структура такой матрицы в дальнейшем может быть представлена в виде графа. Для рассматриваемой выборки граф явно перегружен и здесь не приводится, хотя отмечаются его следующие особенности. Все элементы графа связанные, за исключением коэффициента густоты речной сети для приводораздельного пояса. Этот коэффициент полностью выпадает из рассматриваемой системы как признак. Граф не является полным, он разорван на отдельные цепи, представляющие соединенные ребрами вершины. Довольно отчетливо выделяются цепи в виде связанных между собой характеристик:

$$KX - Z;\ I_p - I_б - \varphi;\ L - N;\ F_п/F - F_c/F;\ H_c/H - H_д/H;$$

$$\alpha - \gamma - M_{п.с} - K_{п.п} - K_{п.с}.$$

В свою очередь граф разбивается на систему подграфов, выделение которых на данном уровне связей весьма проблематично.

Коэффициентам поясности, максимально возможному испарению, невязке водного баланса и площади водосборов рек, если их рассматривать в виде вершин графа, свойственны минимальные нагрузки связей. Первые характеристики из этой группы признаков связаны с другими преимущественно отрицательно.

Такой характер связей, возможно, обусловлен спецификой расчета коэффициентов поясности как отношения частного к общему. Низкая величина выборочного критического коэффициента корреляции (0,28) создает перегруженность графа, и из-за этого трудно выделить подграфы, что необходимо для выделения групп взаимосвязанных признаков. Преодоление перегруженности графа возможно с помощью перехода к более высокому значению выборочного коэффициента корреляции. При анализе групп признаков в качестве сильных связей (как положительных, так и отрицательных) рекомендуется использовать значения коэффициентов корреляции в интервале от 0,7 до 1,0 /Крамбейн, Грейбилл, 1969/. Если значения коэффициентов корреляции заключены в интервале от 0,4 до 0,7, связь считается умеренной. Значения корреляционных коэффициентов менее 0,4 говорят об отсутствии связей или о том, что они очень слабы. На основе этих принципов на рис. 5 построен граф, в котором показаны связи, значения коэффициентов корреляции которых превышают величину 0,4. Сильные связи обозначены жирной линией.

На данном уровне связей между признаками из системы совокупности дополнительно выпадают такие признаки, как величина невязки водного баланса, коэффициенты уклонов долинного пояса и густоты речной сети склоновых и долинных поясов. Общая закономерность коэффициентов поясности в рассматриваемой совокупности признаков – слабый информационный вклад. Эти признаки формируют преимущественно умеренные и слабые связи с другими, но могут быть полезны при рассмотрении морфоструктурных особенностей территории. Использование коэффициентов поясности на основе современных сведений о поясной структуре формирования ресурсов подземных вод в горно-складчатых областях требует выражения через более информативные показатели.

Не представительны по тем же соображениям такие характеристики, как длины порядков магистральных водотоков (l_n), залесенность речных водосборов (C), средние отметки речных русел (H_p), объем зоны дренирования ($V_{з.д}$), суммарная длина речной сети в бассейне (l_o), расстояние до наиболее удаленной точки водосбора (L), его площадь (F).

Из морфометрических признаков более перспективно использование таких, как частота (S) и густота речных потоков (D), возможность эрозионного вреза рек ($h_{вр}$) и комплексный показатель расчленения рельефа (K). Данные признаки являются (частично) производными от перечисленных ранее и обладают в рассматриваемой совокупности большим количеством связей на более высоком уровне.

Использование на первом этапе факторного и корреляционного анализов позволило отобрать из совокупности 37 признаков 14, связанных в системе с характеристиками подземного стока. Повторный расчет по алгоритмам позволил выделить более тонкую структуру взаимосвязанных процессов формирования ресурсов подземных вод и дать основу для последующего районирования по комплексу формирующих признаков.

В результате расчетов было выделено 12 факторов, которые исчерпали 99,7% общей дисперсии признакового пространства. Наибольший интерес с точки зрения интерпретации факторной матрицы представляют пять первых факторов, вобравших 90% общей дисперсии. Следуя выражению (3.8), количество значимых факторов может быть сведено до трех.

Первый имеет преимущественно высокие факторные нагрузки на все показатели, за исключением средней высоты водосбора, плотности тектонических нарушений, мощности эрозионного вреза рек, их уклонов и частоты речных потоков. Для этих признаков свойственны умеренные или приближающиеся к силь-

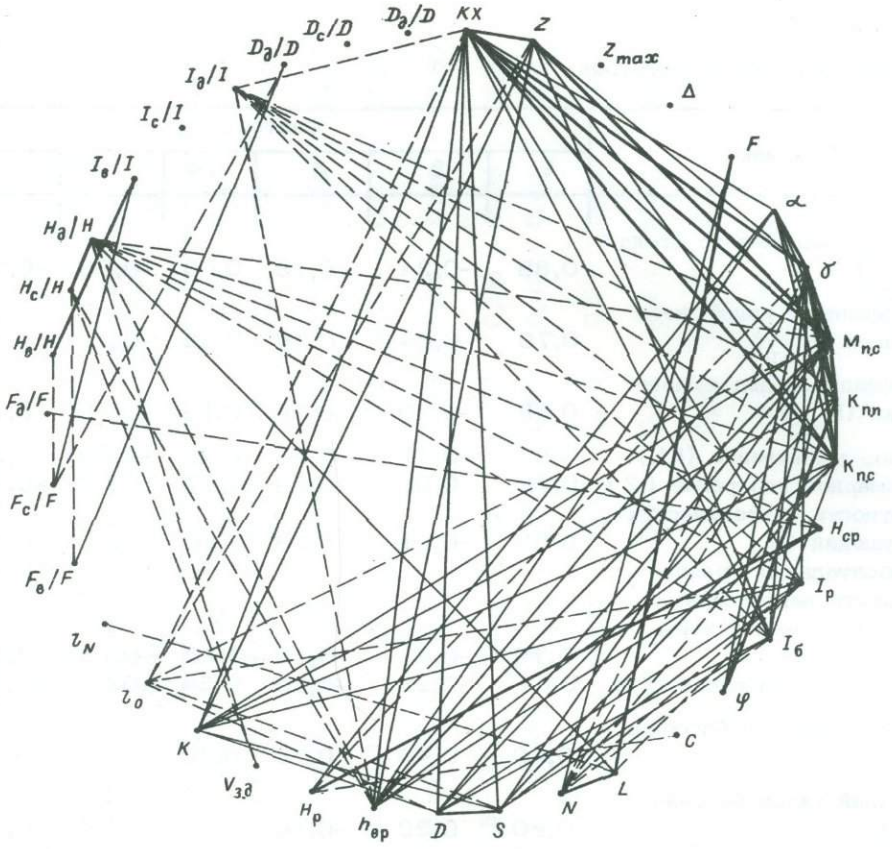


Рис. 5. Граф связей в системе признаков (величина коэффициента корреляции больше 0,4).

ным нагрузки. Следовательно, данный фактор на первом этапе можно интерпретировать как наиболее общий, отражающий основные черты формирования ресурсов подземных вод через комплекс взаимосвязанных признаков (табл. 5).

Второй фактор не имеет сильных нагрузок, но достаточно высокие положительные они выявлены для показателей частоты и густоты речной сети, для них же и в первом факторе свойственны высокие факторные нагрузки. Интересен факт высокой отрицательной нагрузки во втором факторе для плотности тектонических нарушений, поскольку заложение речной сети в горно-складчатых областях происходит в основном по зонам крупных тектонических разломов, и следовательно, речная сеть в региональном масштабе должна соответствовать этим нарушениям. Объединение признаков в факторе с различными знаками, возможно, произошло вследствие того, что расчет плотности тектонических нарушений приводился к топографическим картам масштаба 1:200 000. Показатели частоты и густоты речных потоков рассчитывались для замыкающего створа. Учитывая, что во второй фактор с умеренными нагрузками входят два таких родственных признака, как густота речной сети и частота речных потоков, этот фактор можно интерпретировать как частично отражающий горизонтальное расчленение рельефа.

Третий фактор имеет сильные факторные нагрузки на признаки, отражаю-

Таблица 5

Матрица факторных нагрузок

Признак	Фактор					
	1	2	3	4	5	6
Модуль подземного стока ($M_{п.с}$)	0,88	-0,24	-0,15	0,03	-0,08	-0,32
Коэффициент подземного питания ($K_{п.п}$)	0,70	-0,34	-0,36	0,34	-0,29	0,08
Коэффициент подземного стока ($K_{п.с}$)	0,88	-0,30	-0,16	0,18	-0,18	0,14
Общее увлажнение (КХ)	0,87	-0,11	0,13	-0,28	-0,09	-0,25
Суммарное испарение (Z)	0,78	0,08	0,12	-0,47	0,07	-0,22
Плотность тектонических нарушений (α)	0,55	-0,61	0,18	0,10	0,30	0,36
Относительный градиент скорости вертикальных движений земной коры (γ)	0,79	-0,41	-0,16	-0,10	-0,13	0,12
Средний уклон русла ($I_р$)	0,67	0,26	-0,15	0,54	0,32	-0,03
Средняя высота бассейна ($H_{ср}$)	0,40	0,40	0,70	0,30	-0,06	0,14
Средний уклон бассейна ($I_б$)	0,80	0,22	-0,10	0,07	0,44	0,0
Частота потоков (S)	0,67	0,55	-0,32	-0,26	-0,14	0,15
Густота речной сети (Д)	0,71	0,55	-0,30	-0,08	-0,10	0,21
Мощность эрозионного вреза рек ($h_{вр}$)	0,61	-0,18	0,71	0,04	-0,16	0,15
Комплексный показатель расчленения рельефа (К)	0,84	0,30	0,30	-0,01	-0,14	0,21
Собственные значения (λ)	7,63	1,83	1,59	0,92	0,60	0,53
Кумулятивный вклад факторов в общую дисперсию в %	54,5	67,6	79,0	85,6	89,9	93,7

щие в основном высотное положение речных систем (среднюю высоту водосборов) и мощность эрозионного вреза рек. Вполне естественно, что эти признаки объединились в один фактор, поскольку последний рассчитывался как разность между первой характеристикой и средней отметкой русла.

В современной литературе /Афанасьев, 1976/ считается, что для горно-складчатых областей одним из ведущих условий формирования подземных вод выступает высотное положение гидрогеологических систем. Анализ факторной и корреляционной матриц (табл. 5, 6) показал отсутствие прямых связей основных параметров подземного стока и высотного положения речных бас-

Таблица 6

Корреляционная матрица признаков

Признак	$K_{п.п}$	$K_{п.с}$	KX	Z	α	γ	$H_{ср}$	$I_{р}$	$I_{б}$	S	D	$h_{вр}$	K
1. $M_{п.с}$	0,74	0,95	0,83	0,67	0,48	0,74	0,20	0,55	0,62	0,47	0,48	0,44	0,58
2. $K_{п.п}$	1	0,87	0,45	0,32	0,43	0,72	0,01	0,50	0,46	0,35	0,42	0,31	0,44
3. $K_{п.с}$		1	0,75	0,57	0,54	0,78	0,20	0,57	0,59	0,45	0,50	0,48	0,60
4. KX			1	0,82	0,54	0,71	0,35	0,42	0,66	0,52	0,50	0,57	0,68
5. Z				1	0,40	0,58	0,31	0,34	0,61	0,59	0,54	0,49	0,67
6. α					1	0,65	0,03	0,21	0,37	0,03	0,08	0,56	0,35
7. γ						1	0,03	0,40	0,53	0,40	0,40	0,43	0,52
8. $H_{ср}$							1	0,39	0,32	0,19	0,26	0,63	0,61
9. $I_{р}$								1	0,71	0,46	0,58	0,22	0,54
10. $I_{б}$									1	0,60	0,65	0,31	0,66
11. S										1	0,93	0,12	0,67
12. D											1	0,15	0,72
13. $h_{вр}$												1	0,74
14. K													1

сейнов. Можно предположить, что влияние высоты местности на формирование подземных вод в большей степени передается через другие признаки (экспозиция склонов и речных бассейнов относительно влагоносных ветров, а следовательно, и величина общего увлажнения; наличие достаточной степени раздробленности горных пород, которая в свою очередь - результат повышенной сейсмичности; площади распространения многолетнемерзлых пород по мере увеличения абсолютных отметок местности и др.).

Четвертый и пятый факторы содержат преимущественно умеренные и слабые нагрузки уклонов речных водосборов и русел рек.

Последующие факторы имеют очень низкие факторные нагрузки, и их вклад в общую дисперсию признакового пространства около 6%, поэтому дальнейшее их рассмотрение не имеет смысла.

3.2.1. Разделение признаков на группы методом вращения факторных осей

В практике использования факторного анализа зачастую трудно (особенно при большом количестве признаков) провести интерпретацию факторной матрицы. Затруднения возникают также при высокой взаимокоррелируемости признакового пространства, когда факторам довольно трудно придать какой-либо определенный физический смысл. В этом случае исследователь сталкивается с проблемой анализа групп генетически однородных признаков.

Эти проблемы разрешимы с помощью методов вращения факторных осей в многомерном пространстве для получения простой структуры факторной матрицы. Взяв за основу идею метода главных компонент (направление главных осей соответствует максимальному рассеиванию признакового пространства), некоторые исследователи считают вращение факторных осей нецелесообразным. Однако приводимые ситуации могут иметь место вследствие значительного отклонения ряда признаков от нормального закона распределения, и в случае, когда многомерное признаковое пространство будет отличным от эллипсоидной формы, вращение факторных осей будет не только оправданным, но и необходимым наряду с применением других методов группировок.

Основные критерии вращения факторных осей были разработаны еще в 30-40-е гг. Л. Тэрстоном, суть их сводится к следующим положениям /Йёреског и др., 1980; Лоули, Максвелл, 1967; Jahn, Vahle, 1970; и др./:

1) для каждой строки матрицы факторных нагрузок хотя бы одно значение должно быть равным или близким к нулю;

2) столбцы должны содержать не менее K нулевых элементов, где K - учитываемые факторы;

3) в любой паре факторов должны существовать признаки, факторные нагрузки которых имеют максимальное значение на один фактор и близкие к нулю на другой, а также признаки, имеющие малые нагрузки на оба фактора;

4) в любой паре факторов должно иметься небольшое число признаков, значение отличающихся от нуля.

Одно из основных отличий матрицы факторных нагрузок после вращения от таковой до вращения состоит в том, что в последней для всех или почти всех признаков высокие нагрузки падают на первые факторы, и для всех этих факторов характерны максимальные вклады в общую дисперсию. В матрице факторных нагрузок после вращения нагрузки признаков стремятся к экстремальным значениям с равномерным распределением их по факторам, вклад которых в общую дисперсию распределяется также равномерно (в отличие от матрицы факторных нагрузок до вращения).

В настоящее время существует довольно много способов вращения фак-

торных осей, но наибольшее распространение получил метод варимакс Кайзера. В его основу положен принцип изменения ортогональных факторных осей до тех пор, пока проекции каждого признака (факторные нагрузки) на факторные оси не примут экстремальные значения, т.е. не будут стремиться к нулю либо к 1.

По Йёреского и др. /1980/, дисперсия квадрата нагрузок j-го фактора может быть записана в виде следующего уравнения:

$$S_j^2 = \frac{1}{P}\sum_{i=1}^{p}(b_{i,j}^2)^2 - \frac{1}{P^2}\left(\sum_{i=1}^{p} b_{i,j}\right)^2, \qquad (3.14)$$

где P – число факторов; $b_{i,j}$ – нагрузка i-го признака на этот фактор. Критерий простоты удовлетворяется после максимизации сумм дисперсий по всем факторам $S_t^2 = \sum_{P=1}^{k} S_j^2$ и нормализации каждой строки матрицы нагрузок путем деления каждого элемента на квадратный корень из соответствующей ему общности (h_i). Общность признака определяется путем суммирования квадратов факторных нагрузок по строкам матрицы факторов, и после вращения строки этой матрицы вновь приводятся к исходному размеру. Окончательный вид максимизируемой величины для матрицы на основе метода варимакс выражается следующей формулой:

$$S_v^2 = P\sum_{j=1}^{k}\sum_{i=1}^{p}(b_{i,j}/h_i)^4 - \sum_{j=1}^{k}\left(\sum_{i=1}^{p} b_{i,j}/h_i^2\right)^2 \qquad (3.15)$$

Вращение факторных осей обычно производится итеративным методом попарно для двух осей до тех пор, пока уменьшение дисперсий нагрузок не достигнет какой-то наперед заданной величины угла поворота. Следует отметить, что при использовании различных процедур вращения, особенно при работе с коррелируемыми признаками, довольно трудно удовлетворить все критерии Терстона. Тем не менее этот этап построения факторных моделей представляется весьма полезным, особенно при анализе исходной информации и оценки адекватности моделей. В табл. 7 представлена матрица факторных нагрузок, обработанная методом варимакс Кайзера.

Максимальные нагрузки на первом факторе характерны для основных параметров подземного стока: модуля, коэффициентов подземного стока и питания, два последних из них – производные от модуля подземного стока и величин общей увлажненности и соответственно речного стока. Этим признакам для Байкальской горно-складчатой области свойственны общие черты пространственной изменчивости, как и основному показателю естественных ресурсов подземных вод – модулю. Обращаясь к корреляционной матрице (см. табл.6), можно видеть, что перечисленные признаки между собой имеют достаточно сильные связи (r = 0,74-0,95). Следовательно, первый фактор можно определить как фактор подземного стока. Второй содержит высокие нагрузки для частоты речных потоков и густоты речной сети, между которыми коэффициент корреляции равен 0,93. Это близкие по методике определения признаки, и в принципе они характеризуют горизонтальное расчленение рельефа речной сетью; этим и определяется название второго фактора. Умеренные факторные нагрузки на второй и последующий факторы имеет комплексный показатель расчлененности рельефа, что подтверждает его комплексное назначение (учет горизонтальной и вертикальной расчлененности).

Таблица 7

Матрица факторных нагрузок после вращения

Признак	Фактор							Общность (h)
	1	2	3	4	5	6	7	
$M_{п.с}$	0,76	0,20	0,12	-0,04	-0,16	0,18	0,02	0,866
$K_{п.п}$	0,92	0,19	0,00	0,00	-0,16	0,15	0,09	0,931
$K_{п.с}$	0,86	0,22	0,16	-0,25	-0,23	0,18	0,05	0,961
KX	0,39	0,27	0,27	-0,60	-0,29	0,08	0,14	0,648
Z	0,21	0,36	0,21	-0,85	-0,18	0,05	0,08	0,975
$H_{ср}$	-0,03	0,11	0,96	-0,09	0,07	0,19	-0,02	0,993
$I_{р}$	0,32	0,30	0,20	-0,05	-0,05	0,86	0,03	0,977
$I_{б}$	0,26	0,40	0,16	-0,27	-0,18	0,38	0,06	0,453
S	0,17	0,92	0,04	-0,24	0,05	0,11	0,07	0,972
D	0,22	0,91	0,11	-0,14	0,00	0,22	0,02	0,967
$h_{вр}$	0,24	0,00	0,74	-0,22	-0,41	-0,04	0,06	0,823
K	0,26	0,56	0,58	-0,25	-0,18	0,13	0,10	0,856
α	0,20	-0,04	0,08	-0,10	-0,92	0,05	0,09	0,964
γ	0,59	0,20	0,02	-0,28	-0,10	0,09	0,59	0,971
Вклад факторов в общую дисперсию в %	22,7	18,8	15,1	12,1	10,7	8,2	6,1	

Третий фактор может быть назван фактором вертикального расчленения рельефа, поскольку с высокими факторными нагрузками в него входят такие признаки, как средняя высота водосборов, мощность эрозионного вреза рек и вышеприведенный показатель. Коэффициент корреляции между этими признаками изменяется от 0,61 до 0,74. Интересно отметить, что комплексный показатель расчленения рельефа, по-видимому, наследует в большей степени черты изменчивости вертикальной расчлененности рельефа, чем горизонтальной, а мощность эрозионного вреза рек для Байкальской рифтовой зоны и сопредельной территории контролируется высотным положением речных бассейнов.

Для четвертого фактора максимальные нагрузки с отрицательными знаками свойственны метеорологическим признакам (общему увлажнению территории и испарению); пятый фактор также отрицательно определяется плотностью тектонических нарушений, а шестой - положительно уклонами речной сети. Уклоны речных водосборов имеют отрицательную нагрузку 0,70 лишь в восьмом, а градиенты вертикальных неотектонических движений распределились с одинаковыми значениями факторных нагрузок 0,59 на первый и седьмой факторы*.

*Нагрузки восьмого и последующих факторов не приводятся из-за громоздкости.

Всего шестью факторами после вращения факторной матрицы охвачено около 86% общей дисперсии, а по признакам - от 46 до 99%.

В данном случае показана принципиальная возможность использования метода вращения для целей классификации признакового пространства. Следует отметить, что наряду с ортогональным вращением факторных осей, рассмотренным выше, существуют различные схемы косоугольного вращения, и иногда, особенно для зависимых признаков, эта разновидность вращения лучшим образом отражает факторную структуру исходных признаков.

3.2.2. Районирование территории по условиям формирования ресурсов подземных вод на основе R-схемы факторного анализа

Как справедливо отмечено в работе И.Н. Елисеевой и В.О. Руковишникова /1977/, информативность признака во многом определяется его связанностью с другими характеристиками, а классификация в пространстве наиболее информативных признаков не может проводиться без оценок их статистических связей. Именно таким оценкам связей и группировкам признаков посвящены предыдущие разделы главы.

Ранее /Хаустов, 1978, 1979б Хаустов, Чесалов, 1981; Писарский, Хаустов, 1982; Хаустов, 1984а/ показана принципиальная возможность использования методов факторного анализа для гидрогеологического районирования по условиям формирования подземного стока.

К достоинствам применения эвристических методов, в частности факторного и компонентного анализов, для целей классификации (районирования) отнесем возможность обходиться без использования мер сходства классифицируемых объектов. Из анализа отечественной и зарубежной литературы в качестве мер сходства предложено более 20 разновидностей аналитических выражений, применение которых, как правило, требует соблюдения ряда условий: максимизации межклассового расстояния, взаиморасположения классов относительно друг друга, задания критериев, на основании которых производится объединение объектов в классы, и др. Выполнение указанных требований не всегда возможно и приводит зачастую к многовариантности схем районирования.

Если возьмем сумму произведений факторных нагрузок на стандартизированные значения признаков по каждому наблюдению, получим значение проекций наблюдений на факторы (3.10). Расположение точек проекций наблюдений в координатах факторов может явиться основой классификации. Однородным классам многомерных наблюдений будут соответствовать относительно изолированные компактные группы точек в факторном пространстве. Дальнейшая процедура классификации сводится к разграничению групп точек, которое может проводиться как с использованием формальных методов, так и визуально. На практике часто оказывается достаточным построение проекций многомерных наблюдений в плоскостях первых главных факторов. Число последних, как правило, значительно меньше размерности признакового пространства. На эти же факторы приходится большая часть дисперсии признаков исходной совокупности. Наиболее информативные признаки для классификации будут иметь максимальные факторные нагрузки в первых факторах, и, чем больше вклад этих факторов в суммарную дисперсию признаков, тем конкретней обособляются в классы объекты.

По мере снижения вкладов первых факторов в общую дисперсию, а следовательно, и уменьшения нагрузок признаков на первые факторы межклассовые расстояния становятся менее явными и возникают определенные трудности при проведении однозначных границ между группировками. В этом случае мы име-

ем дело с размытыми множествами, и для преодоления этих неопределенностей необходимо проведение дополнительных операций.

Опыт работы с R-схемой факторного анализа для целей районирования показал, что наиболее эффективно использование первых факторов, поскольку они в силу специфики метода характеризуют процессы значительных пространственных масштабов. Факторы с большими номерами дают представление о мелкомасштабных (локальных) процессах.

На рис. 6 представлена диаграмма распределения проекций многомерных наблюдений на 1-й и 2-й факторы исходной совокупности объектов - 80 речных бассейнов по 14 признакам (см. табл. 5). На первом этапе все множество точек можно разделить на три подмножества: A_1, A_2 и Б. Территориально подмножества A_1 и A_2 включают бассейны рек, преимущественно формирующие подземный сток в пределах горных хребтов, обрамляющих котловину оз. Байкал (за исключением рек на участке Бол. Голоустная - Сарма). К этим подмножествам относятся реки Чикойской горно-складчатой области.

Экстремальные проекции многомерных наблюдений свойственны рекам Южного Прибайкалья и левобережным притокам р. Баргузин. Это подмножество образует самостоятельную таксономическую единицу A_2 в пределах всего множества А. Подмножество Б, имеющее отрицательные проекции наблюдений на первый фактор, представлено речными бассейнами Западного Забайкалья.

С целью более четкого разграничения точек на изолированные группы, а также получения более дробной схемы районирования для подмножеств A_1 и Б был произведен раздельный расчет по той же схеме факторного анализа. В результате таких расчетов для отдельных подмножеств, состоящих из 25 и 50 членов, произошло некоторое ослабление факторных нагрузок и корреляционных связей между признаками (табл. 8, 9, 10). Достаточно высокие факторные нагрузки в подмножестве A_1 остаются для модулей и коэффициентов подземного стока и общего увлажнения. Заметно снизились нагрузки для комплекса морфометрических и геолого-структурных показателей, а также коэффициента подземного питания. Снижение, возможно, произошло вследствие исключения из расчетов подмножества A_2, поскольку группа речных бассейнов имеет экстремальные значения большинства признаков. Высокие отрицательные нагрузки на второй фактор в подмножестве A_2 имеют уклоны рек и показатели горизонтального расчленения рельефа (частота и густота речной сети). В отличие от факторной матрицы для всей совокупности бассейнов этого подмножества произошла довольно существенная перестройка факторной

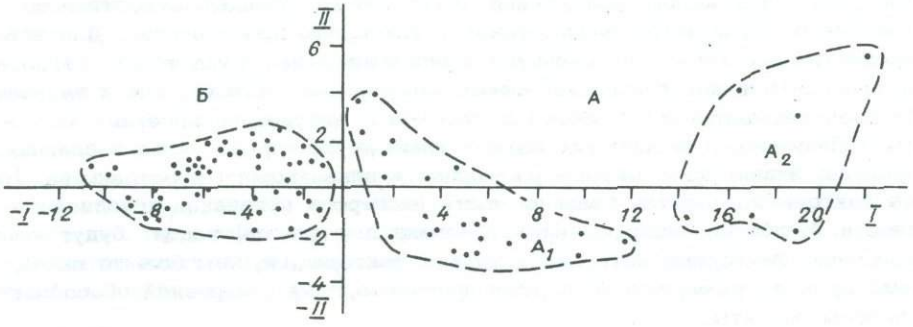


Рис. 6. Диаграмма распределения проекций многомерных наблюдений в плоскости двух факторов.

Таблица 8

Матрица факторных нагрузок для подмножеств A_1 и Б

Признак	Подмножество A_1, n=25 объектов				Подмножество Б, n= 50 объектов			
	Фактор				Фактор			
	1	2	3	4	1	2	3	4
$M_{п.с}$	0,83	-0,25	-0,26	-0,21	0,85	-0,39	-0,03	-0,21
$K_{п.п}$	0,35	-0,45	-0,69	0,30	0,53	-0,68	0,04	-0,07
$K_{п.с}$	0,74	-0,38	-0,43	0,09	0,80	-0,47	-0,01	-0,24
KX	0,62	0,26	0,20	-0,64	0,83	0,16	0,16	0,13
Z	0,29	0,39	0,28	-0,70	0,58	0,44	-0,04	0,31
$H_{ср}$	0,36	-0,36	0,79	0,20	0,49	0,62	-0,40	-0,08
$I_р$	0,45	-0,74	0,05	0,05	0,16	0,33	0,58	0,08
$I_б$	0,54	-0,42	0,07	0,31	0,56	-0,04	0,48	0,50
S	-0,56	-0,63	0,02	-0,37	0,64	0,40	0,39	0,20
D	-0,36	-0,83	0,05	-0,24	0,51	0,42	0,52	-0,30
$h_{вр}$	0,56	0,24	0,65	0,36	0,68	0,04	-0,59	0,28
K	0,46	-0,32	0,70	0,23	0,81	0,32	-0,26	-0,05
α	0,56	0,58	-0,18	0,16	0,36	-0,42	0,21	0,65
γ	0,58	0,20	-0,56	0,05	0,47	-0,58	0,07	-0,23
Собственные значения (λ)	4,07	3,12	2,75	1,59	5,41	2,53	1,65	1,20
Суммарный вклад факторов в общую дисперсию	29,1	51,3	71,0	82,3	38,7	56,7	68,5	77,1

структуры признаков и распределения вкладов факторов в общую дисперсию. Так, например, величины испарения имеют сильную факторную нагрузку лишь в четвертом факторе, а коэффициент подземного питания - в третьем. Интерпретация этого фактора интересна тем, что он сохраняет достаточно высокий вклад в общую дисперсию и вскрывает дополнительные особенности структурных связей между признаками в Байкальской рифтовой зоне.

Так, высокая отрицательная нагрузка для коэффициента подземного питания рек и высокие положительные нагрузки для средней высоты речных бассейнов, мощности эрозионного вреза рек и комплексного показателя расчленения рельефа указывают, что по мере повышения высоты местности происходит снижение роли подземного стока в общем речном, несмотря на увеличение роли горизонтального расчленения рельефа. В пределах региона этот факт проявляется не столь контрастно.

При сравнении корреляционных матриц (см. табл. 9) также отмечается ослабление связей по всем признакам, а результаты вращения матрицы факторных нагрузок подмножества A_1 не противоречат результатам аналогичной про-

Таблица 9

Матрица парных коэффициентов корреляции для подмножества A_1

Признак	$M_{п.с}$	$K_{п.п}$	$K_{п.с}$	KX	Z	α	γ	$H_{ср}$	$I_р$	$I_б$	S	D	$h_{вр}$	K
1. $M_{п.с}$	1	0,53	0,92	0,57	0,21	0,26	0,49	0,14	0,41	0,46	-0,19	-0,05	0,21	0,24
2. $K_{п.п}$		1	0,79	-0,25	-0,37	-0,01	0,50	-0,21	0,40	0,15	-0,05	0,12	-0,21	-0,05
3. $K_{п.с}$			1	0,23	-0,09	0,22	0,52	0,07	0,49	0,37	-0,18	0,03	0,13	0,22
4. KX				1	0,68	0,38	0,28	0,16	0,06	0,38	-0,24	-0,26	0,32	0,17
5. Z					1	0,10	0,19	0,01	-0,23	0,13	-0,14	-0,24	0,22	0,14
6. α						1	0,51	-0,16	-0,03	0,08	-0,69	-0,65	0,36	-0,06
7. γ							1	-0,31	0,10	0,16	-0,35	0,34	0,03	-0,10
8. $H_{ср}$								1	0,49	0,40	-0,06	0,11	0,66	0,83
9. $I_р$									1	0,69	0,15	0,42	0,05	0,43
10. $I_б$										1	-0,09	0,15	-0,02	0,20
11. S											1	0,88	-0,49	-0,05
12. Д												1	-0,42	0,16
13. $h_{вр}$													1	0,79
14. K														1

Таблица 10

Матрица парных коэффициентов корреляции для подмножества Б

Признак	$M_{п.с}$	$K_{п.п}$	$K_{п.с}$	KX	Z	α	γ	$H_{ср}$	$I_{р}$	$I_{б}$	S	D	$h_{вр}$	K
1. $M_{п.с}$	1													
2. $K_{п.п}$	0,72	1												
3. $K_{п.с}$	0,98	0,80	1											
4. KX	0,72	0,16	0,60	1										
5. Z	0,28	-0,03	0,20	0,55	1									
6. α	0,29	0,33	0,27	0,19	0,19	1								
7. γ	0,62	0,46	0,61	0,39	0,02	0,38	1							
8. $H_{ср}$	0,24	-0,17	0,14	0,58	0,47	-0,19	-0,12	1						
9. $I_{р}$	0,08	-0,05	0,05	0,02	0,20	-0,03	-0,09	0,23	1					
10. $I_{б}$	0,38	0,33	0,36	0,32	0,32	0,52	0,14	0,07	0,35	1				
11. S	0,36	0,08	0,29	0,60	0,50	-0,07	0,24	0,32	0,23	0,78	1			
12. D	0,26	0,08	0,26	0,37	0,30	-0,02	0,04	0,26	0,33	0,34	0,78	1		
13. $h_{вр}$	0,49	0,36	0,46	0,54	0,45	0,24	0,17	0,57	-0,14	0,24	0,16	0,00	1	
14. K	0,51	0,27	0,47	0,63	0,52	0,12	0,10	0,64	0,05	0,33	0,53	0,53	0,82	1

цедуры для совокупности объектов. В генетические группы объединились те же характеристики, за исключением дополнительного объединения в одну группу уклонов водосборов и речных русел. Это свидетельствует о том, что в Байкальской рифтовой зоне указанные признаки имеют одинаковую пространственную изменчивость в то время, как в целом для региона она не характерна.

Для подмножества Б также произошло снижение факторных нагрузок по всем признакам (см. табл. 8), но в отличие от подмножества A_1 факторная структура сохранилась почти прежней, хотя заметно (как и в подмножестве A_1) снизились вклады факторов в общую дисперсию признаков. Корреляционные связи между признаками также заметно уменьшились, но не столь контрастно, как в подмножестве A_1.

Сравнительный анализ факторных матриц подмножеств A_1 и Б показывает, что при таком разделении всей совокупности объектов факторные нагрузки для модулей и коэффициентов подземного стока остались примерно соизмеримыми и снизились по сравнению с предыдущим расчетом незначительно, а коэффициент подземного питания заметно снизил свою нагрузку для обоих подмножеств. Последняя характеристика для подмножества A_1 имеет более высокие коэффициенты корреляции с градиентом вертикальных неотектонических движений и уклонами речных русел, чем в подмножестве Б, но в отличие от последнего теснее связана с модулем подземного стока (см. табл. 10).

Модуль подземного стока для первого подмножества на уровне умеренных значений коэффициентов корреляции связан с такими характеристиками, как увлажнение, градиент вертикальных движений земной коры, уклоны рек и водосборов. В подмножестве Б связь этого признака с увлажнением проявляется сильнее и достигает значения 0,72, с градиентом выше – 0,62, а с уклонами становится слабой, но зато увеличивается с показателем мощности эрозионного вреза рек и комплексным показателем расчленения рельефа. Аналогичная структура связей выявляется в подмножествах для коэффициента подземного стока.

Обращаясь вновь к матрицам факторных нагрузок (см. табл. 8), можно видеть, что для величин увлажнения в подмножестве A_1 заметно снизилась нагрузка по сравнению с подмножеством Б. Интересно также отметить факт увеличения нагрузки для испарения во втором подмножестве по сравнению с первым. Такие закономерности не противоречат общим принципам формирования подземного стока в регионе, поскольку группа речных бассейнов подмножества Б (бассейн р. Селенги) расположена в зоне недостаточного увлажнения и роль испарения в формировании пространственной изменчивости параметров естественных ресурсов подземных вод несравненно больше, чем в Прибайкалье. Аналогичные закономерности выявляются для совокупности признаков, характеризующих расчленение рельефа. Их роль в формировании изменчивости параметров подземного стока в Забайкалье больше по сравнению с Прибайкальем. Для последнего региона вообще характерны мелкомасштабные процессы, оказывающие влияние на пространственно-временную изменчивость подземного стока. Для региона Прибайкалья можно продолжать последующие итерации дальнейшего расчленения структуры корреляционных связей с целью получения более мелкомасштабных эффектов воздействия комплекса признаков на формирование подземного стока.

Проекции многомерных наблюдений на факторные оси для подмножеств A_1 и Б изображены на рис. 7. Для первого подмножества выделились две аномальные точки, принадлежащие бассейнам рек Улюкчикан и Нестериха (левобережные притоки р. Баргузин). Подземный сток этих рек формируется преимущественно в интенсивно раздробленных тектоническими нарушениями, горных породах; по условиям и количественным признакам процессы образования близки

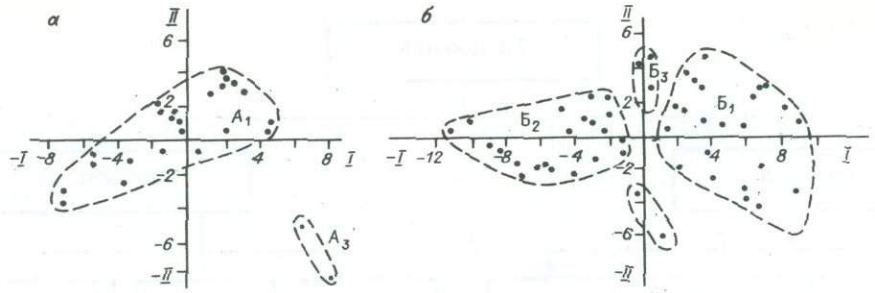


Рис. 7. Проекции многомерных наблюдений на факторные оси для подмножеств $А_1$ (а), Б (б).

к бассейнам рек с распространением карбонатных карстующихся пород (Южное Прибайкалье). Это позволяет объединить точки с указанным районом или считать бассейны рек Улюкчикан и Нестериха за самостоятельный таксон $А_3$. Последнее наиболее предпочтительно, поскольку данный район отличается своеобразной структурой водно-солевого баланса /Хаустов, 1982в/. Остальные точки объединились в относительно компактную группу, хотя, как указывалось выше, возможны дальнейшие итерации этого подмножества. Подмножество Б в свою очередь также распадается на две основные группировки с примерно равным количеством точек (см. рис. 8). Пять промежуточных точек характеризуют переходные между двумя районами условия, которые могут быть представлены так же, как самостоятельные таксоны, либо объединены районом.

Общая схема районирования бассейна оз. Байкал в рамках предлагаемой таксономии (см. табл. 1) распадается на крупные области А и Б (рис. 8). Область А (в территориальном отношении) включает преимущественно бассейны рек горного обрамления оз. Байкал (за исключением центральной части западного побережья), верховья бассейна р. Чикой и часть Хэнтэй-Даурского неотектонического поднятия, расположенного в Забайкалье (рис. 9). Формирование подземного стока здесь происходит в основном в кристаллических массивах и впадинах байкальского типа, которые выполнены хорошо проницаемыми кайнозойскими отложениями. Высокое увлажнение территории обеспечено ориентацией горных хребтов относительно влагоносных масс воздуха. Интенсивная расчлененность рельефа, большие уклоны местности и речной сети способствуют активному водообмену подземных вод во вмещающих породах и взаимодействию поверхностных и подземных вод.

Переходя к более низкому таксономическому уровню в пределах этой области, можно выделить три округа. Округ $А_1$ является репрезентативным для приведенных выше условий формирования подземного стока и включает большинство речных бассейнов области. Округа $А_2$ и $А_3$ характеризуются экстремальными условиями формирования подземного стока для области А и в целом для провинции. Набор признаков для этой совокупности речных бассейнов заметно превышает таковые для округа $А_1$. Подземный сток формируется либо в карбонатных карстующихся породах (реки Южного Прибайкалья), либо в породах, интенсивно разбитых разрывными нарушениями (реки Нестериха, Улюкчикан, Мысовка и др.). Иногда эти условия проявляются наряду с максимальным количеством атмосферных осадков для региона, в результате чего параметры подземного стока приобретают максимальные значения (реки Южного Прибайкалья).

Рис. 8. Иерархическая структура таксономических подразделений.

Область Б расположена в центральной части Селенгинской Даурии. В строении речных бассейнов наряду с гидрогеологическими массивами превалирующее место могут занимать межгорные впадины забайкальского типа, выполненные (*по сравнению с впадинами байкальского типа*) слабопроницаемыми коллекторами. Область находится в условиях аридного и полуаридного климата и может быть подразделена на три округа.

Переходный округ $Б_1$ от областей А и Б приурочен к вторичным подветренным склонам горных хребтов, имеет достаточно благоприятные условия для формирования ресурсов подземных вод, но меньшая увлажненность по сравнению с областью А не позволяет обеспечить высокие параметры. К этому округу отнесены бассейны рек западного побережья оз. Байкал. Реки центральной части этого побережья формируют свой подземный сток под влиянием примыкающего фланга Сибирской платформы в условиях несовпадения поверхностного и подземного водосборов, что приводит к изменению количественных параметров подземного стока в замыкающих гидрометрических створах.

В восточной части Забайкалья округ $Б_1$ охватывает левобережные притоки р. Хилок, непосредственно примыкающие к Чикойскому поднятию, и южную часть Селенгинской Даурии. Этот округ по условиям формирования подземного стока – переходный от округа $Б_2$ к $А_1$.

Округ $Б_2$ занимает центральную часть Селенгинской Даурии. Подземный сток формируется в условиях недостаточного увлажнения при возрастающей роли испарения как с поверхности водосборов, так и непосредственно из зоны аэрации. Дренирующее влияние на водоносные горизонты оказывают такие крупные притоки Селенги, как Хилок, Джида, Чикой и Уда. Большое значение на изменчивость подземного стока как внутри года, так и в многолетнем разрезе приобретает мощность эрозионного вреза рек. Принципиальное отличие формирования ресурсов подземных вод в этом регионе заключается в том, что подавляющая их часть формируется в межгорных впадинах забайкальского типа. Последние представляют систему артезианских бассейнов со своеобразными условиями образования подземных вод, их основные закономерности детально изложены ранее /Пиннекер и др., 1968; Степанов, 1980; Ясько, 1982; и др./.

Округ $Б_3$ представлен аномальными притоками в верховьях р. Уда и в средней части р. Чикой. Признаки, характеризующие основные закономерности формирования подземного стока, и сами параметры имеют здесь минимальные значения, что позволяет считать этот округ наиболее неблагоприятным для формирования естественных ресурсов подземных вод всей провинции.

Рис. 9. Схема районирования бассейна оз. Байкал по условиям формирования естественных ресурсов подземных вод на основе R-факторной модели.

С позиций естественных ресурсов подземных вод в рамках факторной модели для Байкальской горно-складчатой области может быть предложена следующая схема их формирования. Увлажнение территории – потенциальный источник пополнения подземных вод. Проникновению атмосферных осадков и конденсированных вод способствует высокая степень трещиноватости горных пород в гидрогеологических массивах, вызванная высокой сейсмичностью и тектонической активностью в регионе. Особенно контрастно это проявляется для Байкальской рифтовой зоны. Во впадинах байкальского типа активное проникновение атмосферной влаги вызвано слабой метаморфизацией рыхлых кайнозойских отложений, выполняющих эти впадины.

Уклоны речных бассейнов, интенсивная расчлененность рельефа (горизонтальная и вертикальная) предопределяют высокую динамичность подземного стока и активную взаимосвязь с поверхностными водами.

Если вновь обратиться к проблеме снижения величин факторных нагрузок и ослаблению корреляционных связей в подмножествах, выявляется некоторая парадоксальность. Казалось бы, при выделении гомогенного подмножест-

ва связи между признаками, по которым идентифицируются классифицируемые объекты, должны быть более тесными, поскольку по мере упрощения структуры системы ее элементам должен быть присущ элементарный характер взаимодействия. Здесь же, расчленяя множество на более однородные подмножества, мы сталкиваемся с ослаблением связей между признаками при переходе к более низкому иерархическому уровню.

Высокие значения коэффициентов корреляции для всей совокупности объектов обусловлены их расчетом по всей статистической структуре выборки, т.е. при наличии экстремальных значений, имеющих максимальные величины среднеквадратического отклонения. Исключение из выборки последних, возможно, привело к снижению факторных нагрузок в подмножествах и коэффициентов корреляции между признаками. Кроме того, можно предполагать, что по мере снижения иерархической значимости подсистем характер взаимодействия между элементами усложняется, хотя по набору признаков объекты объединяются в группы с более высокими корреляционными связями.

На основании сказанного весьма перспективным представляется создание политомических классификаций, для которых основная методическая особенность заключается в поэтапном построении системы признаков, классифицирующих объекты. Эти признаки могут быть заданными на разных этапах классификации в явном (в виде первичных признаков) и неявном видах (через определенные коэффициенты, отражающие большую часть информации). На каждом этапе классификации необходимо проводить оценку информативной емкости признакового пространства с тем, чтобы иметь возможность рассмотреть структуру связей в выделенной системе (таксоне, классе) и иметь возможность дополнить или сократить признаковое пространство. Такой методический прием позволит выявить более тонкую структуру связей, не только между признаками на разных иерархических уровнях, но и изучить взаимодействие между элементами системы через комплекс связей. Этот метод построения классификаций не противоречит системному принципу исследования природных систем, логике классификации и использован при создании схемы районирования горно-складчатых областей по условиям формирования подземного стока (см. табл. 1).

3.2.3. Методика построения и анализ карт факторных поверхностей*

В практике гидрогеологических исследований довольно широко используется построение различного вида карт с помощью ЭВМ. Методику построения таких карт и некоторые алгоритмы можно найти почти во всех монографиях, посвященных применению математических методов в геологии. Большинство алгоритмов нацелено преимущественно на решение основных видов задач: получение карт, учитывающих особенности отдельных регионов; достижение наибольшей степени сходимости результатов построенных карт с физическими данными; упрощение процедуры построения карт за счет эффективного использования ЭВМ и графопостроителей. Построение универсальных алгоритмов, сочетающих перечисленные направления, – весьма сложная и проблематичная задача, требующая квалифицированного участия математиков-программистов.

Вследствие недостатка или большой вариабельности данных на сравнительно ограниченных площадях построение карт вручную становится невозможным. Традиционно используемые методы аппроксимации или интерполяции по-

*В написании данного раздела принимал участие А.Н. Судаков.

линомами или другими функциями с носителями, совпадающими с областью задания информации, приводят к сильному осциллированию полученного поля и в конечном счете значительно искажают реальную картину распределения признаков в пространстве.

В частности, при использовании полиноминальной интерполяции по методу наименьших квадратов (наиболее распространенный вариант) степень полинома подбирается таким образом, чтобы среднеквадратическая ошибка точечного приближения была минимальна. При этом алгоритм вычисления можно использовать лишь в случае сравнительно простого характера приближающей функции. Использование полиномов высоких степеней приводит к получению с их помощью информации лишь на ограниченных площадях, а возможности экстраполяции информации должны задаваться определенными граничными условиями. Метод наименьших квадратов позволяет достигать минимума погрешности лишь в точках наблюдений при относительно частой и равномерной сети точек.

Для нерегулярной сети наблюдений (на наш взгляд) более перспективен метод непересекающихся треугольников, поскольку выбор их сети производится многократно и при этом учитывается поведение изолиний в соседних точках, что позволяет процедурой построения триангуляционной сети провести в какой-то мере сглаживание случайных ошибок измерений. Трудности могут возникнуть в выборе оптимальной сети разбиения треугольников с помощью ЭВМ. Предлагаемый алгоритм, разработанный и реализованный на ЭВМ БЭСМ-6 А.Н. Судаковым, базируется на идеях метода конечных элементов, теории сплайнов, финитных функций с конечным носителем и позволяет графически отображать информацию о пространстве с двумя независимыми переменными. Основные идеи алгоритма сводятся к следующим процедурам.

Ставится задача (определенная на некотором дискретном множестве $R = \{x, y/i = 1,\ldots,N\}$) провести изолинии двух переменных $z = f(x, y)$. Точки наблюдений принадлежат замкнутой области Q и расположены в ней нерегулярным образом. Для аппроксимации эта область разбивается на систему непересекающихся элементов (треугольников) с вершинами в точках множества R. Система треугольников, имеющих общую точку (вершину), на первом этапе образует конечный носитель для аппроксимирующей функции. Область разбивается на треугольники следующим образом. Выбирается точка (x_o, y_o), имеющая минимальные значения координат $(x_o, y_o) = \min(x_i, y_i)$. Следующая точка (x_1, y_1) принимается таким образом, чтобы все оставшиеся лежали слева от прямой L_o, которая соединяет точки (x_o, y_o) и (x_1, y_1), т.е. удовлетворяет условию

$$\begin{vmatrix} x_o, y_o \\ x_1, y_1 \\ x_i, y_i \end{vmatrix} < 0 \forall_i .$$

Данная точка перемещается во вспомогательный массив K_1, затем вновь выбирается точка, которая должна располагаться слева от линии $L_{1,2}$. Операции повторяются до тех пор, пока последняя точка не совпадет с начальной (x_o, y_o). В случае попадания нескольких точек на прямую $L_{i,i+1}$ они упорядочиваются по возрастанию расстояния от координаты (x_i, y_i) и в этом же порядке перемещаются в множество K_1. После завершения этого этапа в массиве K_1 расположены упорядоченные точки, которые задают на плоскости $O_{x,y}$

выпуклый многоугольник (из множества R эти точки исключены). В множестве R снова выбирается точка с минимальными координатами, и по описанным операциям образуется массив K_2 и т.д. до исчерпания точек в множестве R. После выполнения всех операций на плоскости O_{xy} получится система вложенных друг в друга выпуклых многоугольников, вершины которых хранятся в массивах K_j (n - число вершин в j-м многоугольнике, j = 1,..., m).

Следующий этап - разбивка области на треугольники внутри смежных многоугольников K_j, K_{j+1}. В многоугольнике выбирается произвольная точка А и находится точка с минимальным расстоянием от точки А в многоугольнике K_{j+1}. Далее производится циклическая нумерация точек в многоугольниках, начиная с А для K_j и с В для K_{j+1}. После этих операций образуется четырехугольник из точек А (x_o, y_o), В $(x_o, y_o)^{j+1}$, $(x_1, y_1)^j$, $(x_o, y_o)^{j+1}$ - верхний индекс относится к номеру многоугольника. Если четырехугольник выпуклый, то проводится минимальная диагональ, в противном случае - диагональ, которая не пересекает сторону АВ. После этого образуется треугольник АВС (С - вершина, к которой проведена диагональ из А или В). Фиксируя номера вершин полученного треугольника, пересылаем С в А или В в зависимости от принадлежности С к одному из треугольников. Затем образуется следующий четырехугольник, и операции повторяются до исчерпания точек хотя бы одного из многоугольников. После этого проверяется, исчерпались ли точки другого многоугольника, и если нет, то последовательно образуются треугольники из последней точки одного многоугольника и вершин другого до исчерпания последнего. Всего по алгоритму будет образовано $n_j + n_{j+1}$ треугольников. Такие операции повторяются для всех многоугольников.

В работе У. Кроули /1973/ указывается, что в общем случае проведенная триангуляризация не оптимальна в смысле минимизации априорной погрешности аппроксимации. В целях устранения этого эффекта для оптимизации разбиения многоугольников на треугольники применяется итерационный локальный алгоритм, который переводит выпуклые четырехугольники с максимальной диагональю путем пересмотра всех вершин (точек множества). После работы алгоритма общая длина сети триангуляции минимальна. Итерационный процесс оптимизации триангуляризации сети придает локальному критерию оптимизации глобальный характер, чем устраняется некоторая произвольность в выборе оптимальной сети треугольников. Далее на каждом треугольнике строится финитная базисная функция. Преимущества выбора такого вида базисных функций достаточно аргументированно изложены в работе Г.И. Марчука и В.И. Агошкова /1981/. После этой процедуры проводится кусочно-непрерывное восполнение поля в выпуклой области, заключенной внутри многоугольника K_1.

На этом этапе можно построить карту. Изолинии поля будут представлены отрезками прямых, парабол и другими видами (в зависимости от степени использованных базисных функций на конечном носителе). Предлагаемый алгоритм позволяет также получить восполнение поля до заданной степени гладкости. В этом случае производится пересчет поля по кусочно-непрерывному восполнению на регулярную сетку с заданными пространственными шагами, затем строятся сплайны заданной степени гладкости и с учетом точности полученных измерений по ним вновь рассчитывается поле. Последний этап - обращение к программному комплексу ГРАФОР из математического обеспечения БЭСМ-6, в результате чего вычерчивается заданное поле. Проверка работы предлагаемого алгоритма на различных природных объектах показала его достаточно высокую эффективность /Судаков и др., 1983/.

Необходимость оперирования комплексом признаков для изучения прост-

ранственной изменчивости гидрогеологических процессов назрела сравнительно давно. Подобный подход открывает новые пути в гидрогеологическом картировании, значительно увеличивая при этом информационную емкость карт различного назначения. Именно о таких картах неоднократно упоминалось в работах Н.В. Роговской и В.Н. Кунина /1969/, Н.В. Роговской /1981/. В практике гидрогеологических исследований в настоящее время известно лишь несколько характерных работ, хотя преимущество таких карт по сравнению с моноэлементными очевидно, особенно для выявления региональных закономерностей изучаемого процесса. За примерами можно обратиться к последним работам в области геохимии, где при картировании оперируют не одним каким-либо показателем, а их ассоциациями /Евдокимова, 1978/.

Выше указывалось, что современными алгоритмами предусматривается расчет значений факторов для каждого объекта. Получаемые переменные в результате разложения исходной информации по ортогональным осям на собственные векторы и значения корреляционной матрицы отражают комбинации исходных признаков для каждого объекта. Таким образом, можно сжать большую часть исходной информации до 2-4 главных факторов, имеющих наибольшие дисперсии. Опыт построения таких карт показывает, что они, как правило, отражают наиболее существенные черты и структурные соотношения изучаемых процессов при условии достаточного набора признаков, характеризующих эти процессы. По мере снижения вкладов факторов выявляются остаточные локальные особенности процессов.

При построении карт факторных поверхностей для Байкальской горно-складчатой области использовался сглаженный вариант предлагаемого алгоритма. Высокие положительные значения первого фактора (вклад в общую дисперсию 54%) приурочены к Байкальской рифтовой зоне, за исключением средней части западного побережья озера, а также к северной части Хэнтэй-Даурского неотектонического поднятия (верховье р. Чикой). Для последнего региона положительные величины значений несоизмеримо меньше, чем для рифтовой зоны (рис. 10). Для Забайкалья (Селенгино-Витимский прогиб) свойственны отрицательные значения факторных поверхностей.

Интересно отметить, что нулевая изолиния в плане довольно точно совпадает с границей Байкальской рифтовой зоны. Как это следует из анализа факторной матрицы (см. табл. 5), поверхность первого фактора можно трактовать в виде интегрального показателя формирования подземного стока. Различие в знаках факторной поверхности указывает на разнонаправленность региональных закономерностей процессов. Полученная схема в целом подтверждает основные закономерности, вскрытые при районировании с использованием проекций наблюдений на первые факторы (см. рис. 9).

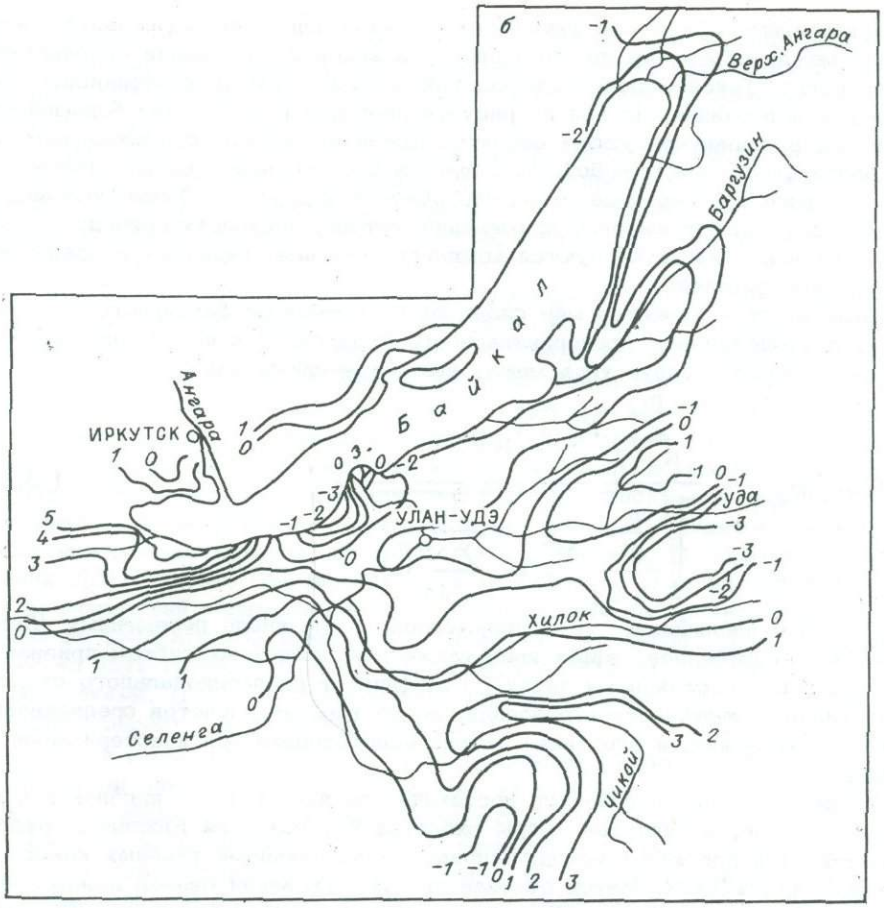
Максимальные положительные значения фактора тяготеют к юго-восточному и северо-восточному побережью озера. Для этих районов свойственны максимальное увлажнение территории, интенсивная горизонтальная и вертикальная расчлененность рельефа, высокая сейсмичность, благоприятствующая раскрытию гидрогеологических систем, а следовательно, и формированию коллекторских свойств водовмещающих пород. Совокупность перечисленных условий выражается через повышенные параметры естественных ресурсов подземных вод. На фоне такой тенденции отмечаются аномалии, выражающиеся в изменении знака фактора на противоположный. Такие аномалии связаны, во-первых, с меньшей увлажненностью на подветренных склонах Приморского хребта (западное побережье озера), во-вторых, для этого района на основе проведенных исследований /Писарский, Русанов, 1981/ установлен факт оттока ресурсов подземных вод к примыкающему флангу Сибирской платформы. Отмеченные природные закономерности сказались на основных параметрах подземного стока для этого района. Их величина минимальна для Прибайкалья.

Рис. 10. Поверхность первого (а), второго (б) факторов (сглаженный вариант).

Положительные (но меньшие по сравнению с Прибайкальем) величины факторной поверхности выявлены для северной части Хэнтэй-Даурского неотектонического поднятия (верховье р. Чикой). Формирование ресурсов подземных вод здесь аналогично Байкальской рифтовой зоне, но меньшая увлажненность территории и менее интенсивная раскрытость гидрогеологических массивов (поднятие более молодое, чем Байкальский рифт) обеспечивают параметры подземного стока несколько выше средних для региона в целом (см. табл. 2).

Область отрицательных значений приурочена преимущественно к мезозойским формам складчатости. Подземный сток формируется в условиях недостаточного увлажнения, и основные черты его формирования близки к платформенным. Относительный тектонический покой территории, широкое развитие межгорных впадин закрытого и полузакрытого типа сказались на прямых показателях подземного стока (они в 5-10 раз ниже, чем в рифте) и на косвенных признаках, характеризующих эти процессы (примерно вдвое ниже). Максимум отрицательных значений фактора показывает наиболее неблагоприятные условия формирования ресурсов подземных вод.

Наибольшая пространственная изменчивость комплекса признаков присуща горному обрамлению котловины озера, причем эта изменчивость интенсив-



но возрастает по мере роста абсолютных отметок местности. Для области отрицательных значений поверхность первого фактора варьирует не столь интенсивно, как для рифтовой зоны.

Поверхность второго фактора (см. рис. 10, б) отражает остаточную часть дисперсии горизонтального расчленения рельефа (около 30%) и с отрицательным знаком плотность тектонических нарушений (около 36%), изменчивость остальных признаков в данном факторе не превышает 20%. Для этой факторной поверхности типичны образование более мелкомасштабных областей изолиний и меньшая вариабельность значений поля.

Третий фактор отражает примерно 50% изменчивости средних высот водосборов и столько же мощности эрозионного вреза рек; изменчивость остальных признаков не превышает 10%, и карта этого фактора здесь не приводится.

3.3. Q-ФАКТОРНАЯ МОДЕЛЬ ФОРМИРОВАНИЯ РЕСУРСОВ ПОДЗЕМНЫХ ВОД

В отличие от R-методов факторного анализа, Q-методы в практике геологических исследований нашли меньшее применение, хотя в зарубежной литературе насчитывается несколько примеров удачного использования этой разновидности факторного анализа.

Применение Q-метода факторного анализа нацелено на описание взаимосвязей между объектами по комплексу признаков на основании определенной меры сходства. Такой мерой сходства, как правило, служат коэффициенты корреляции или ковариации, но для их расчета необходим достаточно большой набор признаков, характеризующих объекты. Признаки должны существенным образом различаться между собой (т.е. быть в достаточной степени независимыми), а также при переходе от одного объекта к другому. Такие условия, по-видимому, сдерживают широкое применение метода, поскольку природные системы чаще всего характеризуются комплексом взаимосвязанных и взаимообусловленных показателей.

Чаще всего в качестве мер сходства в Q-методе факторного анализа используют коэффициенты пропорционального сходства Имбри и Парди, которые выражаются через косинус угла между векторами-объектами

$$\cos \Theta_{i,j} = \frac{\sum_{k=1}^{m} X_{i,k} X_{j,k}}{\sqrt{\sum_{k=1}^{m} X_{i,k}^2 \sum_{k=1}^{m} X_{j,k}^2}}, \qquad (3.16)$$

где $X_{i,k}$ – i-е наблюдение k-й переменной, m – число переменных /Девис, 1977/. Этот коэффициент эффективен также для замкнутых систем признаков (их сумма равна постоянному числу). Коэффициент пропорционального сходства аналогичен коэффициенту корреляции, но не требует расчетов среднеквадратического отклонения, и его также можно использовать при кластеризации признаков.

Когда число число объектов достаточно велико, а число признаков несоизмеримо меньше, в качестве меры сходства используется Евклидово расстояние (метрика) и применим метод, основанный на дуализме главных компонент /Белонин и др., 1982/. Метод основан на существовании связей между преобразованием в пространстве признаков и пространстве наблюдений. Данный метод предлагает одновременное проецирование на факторные оси признакового пространства и пространства наблюдений и позволяет наглядно отразить воздействия тех или иных признаков в пределах группировок.

Итак, в основу Q-метода факторного анализа положена замена признаков наблюдениями (объектами), т.е. транспонирование матрицы $N \times n$ в матрицу $n \times N$, где N – число признаков, а n – число объектов. Первый этап в этом методе – расчет коэффициентов корреляции либо другой меры сходства между объектами. Все дальнейшие операции расчетов идентичны R-схеме факторного анализа и основаны на нахождении собственных значений и собственных векторов корреляционной матрицы, при этом истинная размерность задачи (число извлеченных факторов) не может быть больше числа объектов, т.е. $N < n$.

Для расчетов использовалась та же совокупность 80 речных бассейнов с тем же числом признаков. Исходные данные нормировались и центрировались относительно их дисперсий. В результате разложения корреляционной матрицы объектов на шесть факторов ими было охвачено 100% общей дисперсии. Первый фактор вобрал 95,7%, второй – 2,5, третий – 1,1, четвертый и последующие – 0,4, 0,2 и 0,1% соответственно. Такое явно неравномерное распределение дисперсии по факторам говорит о том, что объекты между собой связаны довольно тесными корреляционными связями по совокупности признаков. И действительно, с помощью корреляционной матрицы можно определить, что диапазон изменения коэффициентов корреляции между объектами по 14 призна-

кам изменяется от 0,79 до 1. Вполне естественно, что первый фактор при таком соотношении исходной информации имеет только положительные факторные нагрузки от 0,89 до 1. Для второго свойственны незначительные положительные и отрицательные нагрузки.

При анализе диаграммы факторных нагрузок в плоскости двух первых факторов выявляется подковообразная форма распределения точек, и в сложившейся ситуации всю совокупность довольно трудно разбить на четкие кластеры, поскольку нельзя выделить ядра классов, а границы последних носят размытый характер.

На рис. 11 отображена попытка сгруппировать исходные объекты с помощью величин их коэффициентов разложения по первому и второму факторам. Из приведенной диаграммы видно, что она также имеет подковообразную форму, но межклассовые расстояния выделяются более четко в сравнении с первым вариантом. О такой форме диаграмм, которые довольно часто встречаются при построении Q-модификации факторных моделей, упомянуто в работе К. Йёрскога и др. /1980/; авторы считают, что своим появлением такая форма распределения точек обязана квадратичной форме зависимости между проекциями наблюдения на две первые оси, что вызвано примерно одинаковыми корреляциями между признаками и, следовательно, объектами. В противовес этому мнению Л.Н. Дуденко (переводчик этой монографии) объясняет появление таких зависимостей полимодальностью распределения выборочной совокупности, а наличие квадратичной зависимости означает, что данная выборка состоит хотя бы из двух групп, в которых существует своя специфика взаимосвязи между признаками. В случае унимодального распределения проекции величин на плоскости образуется выпуклое множество.

Следует отметить, что при наших расчетах коэффициенты корреляции между факторами равны нулю, что указывает на сохранение ортогональности осей. В первом приближении можно выделить несколько довольно компактных группировок. При этом под классом будем понимать наличие хотя бы двух точек минимум в одной группировке, кроме того, для некоторых точек, находящихся в непосредственной близости от выделенных классов, необходимо также учитывать их территориальное соответствие. Из всей совокупности точек девять находятся на значительном расстоянии от центров группировок. Они могут быть интерпретированы как самостоятельные таксономические единицы, и на рис. 11 для этих точек условные обозначения не приводятся.

Если обратиться к предыдущей схеме районирования с помощью R-метода, то можно также видеть, что на всех стадиях процедуры выделяются основные группировки точек: одна из них соответствует Байкальскому рифту, другая - Селенгино-Витимскому прогибу. Положительные значения коэффициентов второго фактора включают реки бассейна Селенги и по сравнению с Прибайкальем характеризуются значительно меньшими величинами подземного стока (0,1 - 2,0 л/(с·км^2)), общего увлажнения территории (350-815 мм), менее расчлененным рельефом (60-480 м/км^2), слабой тектонической раздробленностью и значительной мощностью четвертичных отложений в долинах крупных рек. В качестве экстремальных условий, наименее благоприятных для формирования ресурсов подземных вод, можно привести основные признаки для первого района. Величина увлажнения минимальна для всей провинции (около 400 мм), испарение достигает 370 мм, наименьшая расчлененность рельефа (комплексный показатель) - 63-78 м/км. Отсюда значения модуля подземного стока - 0,08-0,2 л/(с·км^2), коэффициент подземного стока - 1-2, а подземного питания - 5-6%.

Для второй области (Байкальской рифтовой зоны) косвенные показатели формирования ресурсов подземных вод несравненно выше. Величина увлажнения в отдельных районах достигает 1600 мм, расчлененность рельефа - 140-

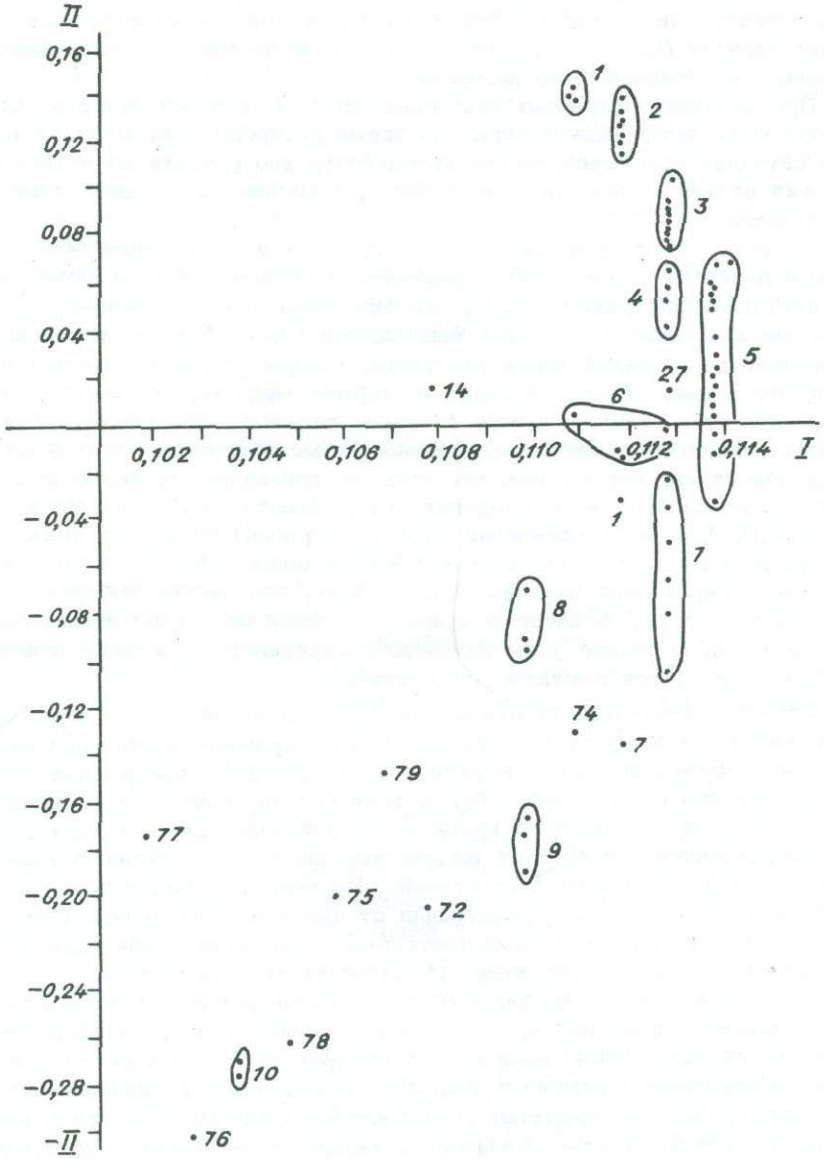


Рис. 11. Диаграмма коэффициентов разложения исходной совокупности объектов в плоскости первых факторов. Цифры у обведенных совокупностей точек соответствуют номерам районов (см. рис. 14); цифры у точек - индивидуальным объектам, не объединившимся в классы.

1100 м/км, плотность тектонических нарушений максимальна в регионе, а величины модуля подземного стока изменяются от 0,9 до 17 л/(с·км^2). В целом для этой области характерна бо́льшая размытость структуры, чем для первой, что свидетельствует о более разнообразных по сравнению с Селенгино-Витимским прогибом условиях формирования подземного стока. В пределах каждой области выделяются более мелкие таксономические подразделения, отличаю-

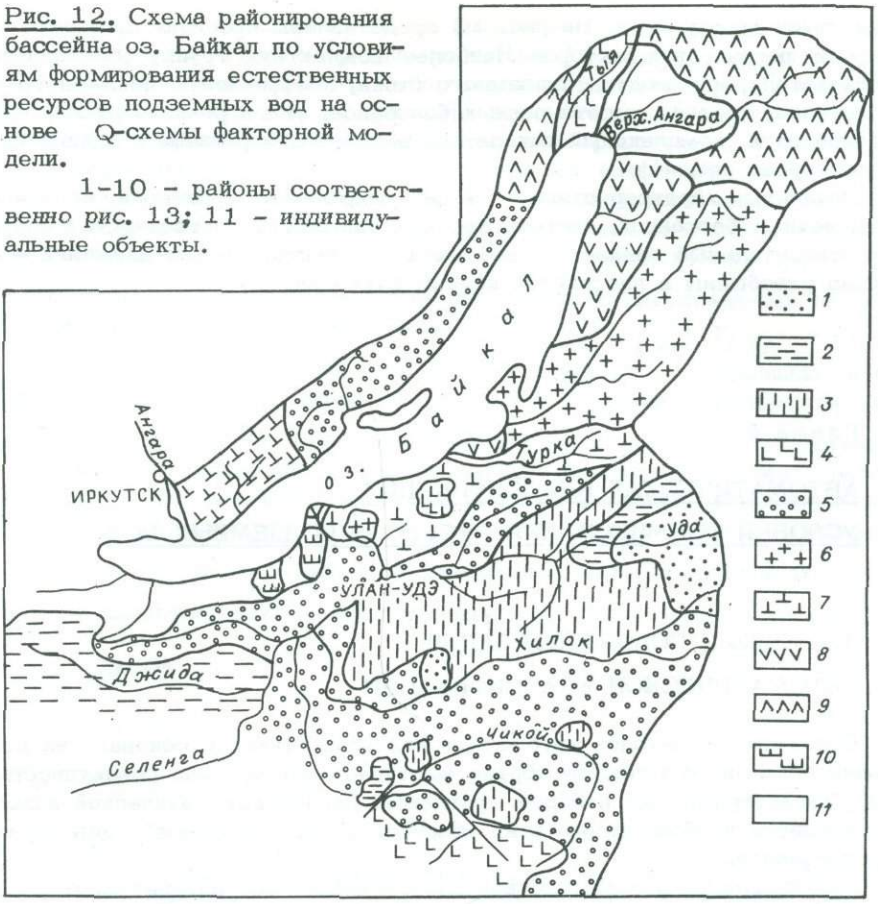


Рис. 12. Схема районирования бассейна оз. Байкал по условиям формирования естественных ресурсов подземных вод на основе Q-схемы факторной модели.

1-10 – районы соответственно рис. 13; 11 – индивидуальные объекты.

щиеся своеобразием формирования ресурсов подземных вод (рис. 12).

Используемый алгоритм Q-схемы факторного анализа не исключает возможности проведения анализа групп признаков, позволяющего выявить внут-

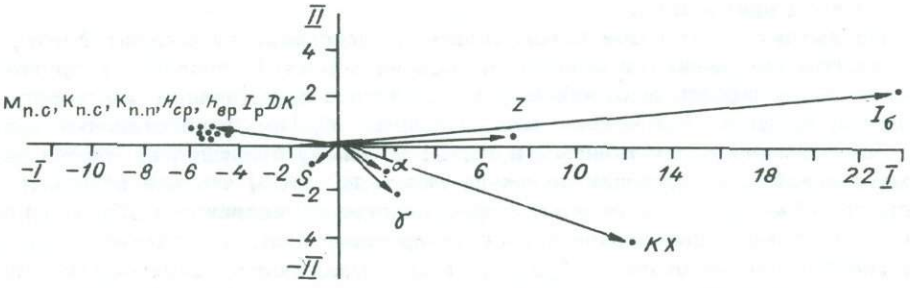


Рис. 13. Проекции признакового пространства в плоскостях первых двух факторов.

ренние связи между ними. На рис. 13 представлены проекции признаков в плоскостях первых двух факторов. Наиболее компактную группу образовали такие показатели, как модуль подземного стока, коэффициенты подземного стока и питания, средняя высота речных бассейнов, уклон речного русла, густота речной сети, комплексный показатель расчленения рельефа и мощность эрозионного вреза рек.

Остальные характеристики – общее увлажнение территории, испарение, уклоны речных бассейнов, частота потоков, плотность тектонических нарушений и относительный показатель вертикальных тектонических движений – значительно разобщены в плоскостях первых факторов.

Глава 4

АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ УСЛОВИЙ ФОРМИРОВАНИЯ РЕСУРСОВ ПОДЗЕМНЫХ ВОД

4.1. ОБЩИЕ ПОЛОЖЕНИЯ ЗАДАЧ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ

Большинство методов автоматической классификации основано на распознавании образов. В качестве образа выступает объект (или совокупность объектов), характеризуемый набором признаков. Задача автоматической классификации сводится к объединению этих объектов в "естественные" типы по комплексу признаков.

При процедуре автоматической классификации все алгоритмы могут быть подразделены на алгоритмы, использующие обучающую выборку с "учителем", и алгоритмы, классифицирующие объекты без обучения. Последняя разновидность классификации по принципам построения считается "объективной", поскольку формальные процедуры классификации перекладываются на ЭВМ, роль исследователя при создании сводится к подготовке исходных данных и интерпретации конечных результатов. Априорные сведения о конечных классах и их числе, как правило, неизвестны, либо по ним имеются самые общие сведения.

В публикациях, посвященных данному вопросу, с различной степенью детальности рассмотрены теория методов, алгоритмы их реализации и примеры практического применения.

Преимущество методов автоматической классификации состоит в том, что в классы объединяются объекты на основе перехода от одного к другому с учетом изменчивости многообразия их свойств с применением некоторого решающего правила. Эти методы самым полным образом удовлетворяют принципам классификации, изложенным в гл. 2. В основу большинства алгоритмов автоматической классификации положена гипотеза о том, что вся исходная совокупность объектов в многомерном пространстве естественным образом распадается на компактные подмножества (в противном случае никакие алгоритмы классификации работать не будут). Под компактностью понимается близость точек, входящих в состав класса, к некоторому эталону (типичному представителю).

Объединение объектов в классы (подмножества) основано на использовании меры сходства (либо функции расстояния) между двумя рассматриваемыми точками. В связи с этим выбор меры сходства (метрики) между объектами - весьма важный этап в процедуре автоматической классификации. Как правило, эта метрика задается в виде какого-то расстояния, в пределах которого объекты попадают в класс, либо с помощью коэффициентов подобия.

В самом общем виде задача автоматической классификации может быть показана следующим образом. Существует конечное множество объектов, которые необходимо классифицировать, $U = \{u_j\}$. Каждый объект $u_j \in U$ обладает m степенями свободы (признаками), дающими представления об основных свойствах, т.е. справедливо выражение $x_j = (x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(m)})$, где x_j - многомерная случайная величина - описана с помощью набора признаков $x_{(1)}, \ldots, x_{(m)}$ либо измерений. Наша задача - объединить множество объектов в таксоны (кластеры, классы, типы) на основе сходства.

Меру сходства объектов в многомерном пространстве можно представить в виде некоторой функции, ставящей в соответствие каждой паре наблюдений (x_i, y_j) *некоторое число* $a_{i,j}$, *характеризующее степень сходства* между объектами u_i и u_j. Основные требования, которые предъявляются к этой мере сходства, заключаются в следующем /Елисеева, Рукавишников, 1977/:

$$a_{i,\,j} \geqslant 0; \tag{4.1}$$

$$a_{i,\,j} = a_{j,\,i}; \tag{4.2}$$

$$a_{i,\,i} \geqslant a_{j,\,j} \quad \forall u_j \in u, \tag{4.3}$$

где u - множество.

Второе равенство приводимых требований предполагает наличие условий симметрии. В противном случае задача классификации становится неопределенной, поскольку она может иметь множество решений, зависящих от порядка рассмотрения объектов множества. Кроме приведенных требований, мера сходства должна характеризовать наикратчайшее расстояние между объектами

$$\forall x_i,\ x_j,\ x_k;\ a(x_{i,\,j}) \leqslant a(x_i \cdot x_k) + a(x_j \cdot x_k). \tag{4.4}$$

Приведенные требования являются весьма жесткими по отношению к классификации естественных объектов, особенно гидрогеологических. Дело в том, что ресурсы подземных вод формируются под влиянием большого числа природных процессов, имеющих различные пространственно-временные масштабы и инерционность. Это несколько затушевывает, размывает границы межклассовых расстояний. Многие процессы довольно трудно выразить через меру и число. Точность же признаков, которые удается измерить, различна и может не удовлетворять решению классификационных задач, а для поиска реальных природных границ естественных процессов требуются значительные усилия.

По сути дела, какие бы методы и вычислительные средства мы ни применяли для классификации природных объектов, мы классифицируем не объекты, а наши знания о них. Приведенные данные указывают на то, что при классификации природных систем, не всегда удается выполнять все требования. Это особенно трудно при выполнении требований симметрии взаиморасположения точек и минимизации расстояния между объектами. Ослабление части приводимых требований приводит к весьма существенному расширению числа функций, используемых для формального выражения однотипности объектов, что, в

свою очередь, отражается на самих классификационных процедурах, качестве и адекватности конечных классификаций.

При создании алгоритмов автоматических классификаций используют в основном три разновидности мер сходства: коэффициенты подобия, которых насчитывается свыше десятка (они применяются в основном в приложении булевых векторов-описаний объектов); коэффициенты корреляции и показатели расстояния. Для некоторых методов (например, кластер-анализа) две последние *разновидности мер применяются поэтапно.*

Использование коэффициентов подобия Рао, Дейка и других эффективно в том случае, когда в качестве признаков используется (вместо эмпирических величин) кодирование информации, т.е. наличие *какого-либо явления обозначается* 1, а его отсутствие - 0. Такие признаки называются дихотомическими. *Примеры подобного построения классификаций для гидрогеологических систем* межгорных впадин Азии достаточно аргументированно показаны в работах К.П. Караванова /1977; 1980/.

Для дихотомических признаков чаще всего в качестве метрики при классификациях используется расстояние по Хеммингу

$$d_{i,j} = \sum_{k=1}^{m} \left| x_i^k - x_j^k \right|. \qquad (4.5)$$

Особенности применения этой метрики здесь не рассматриваются, поскольку используемые при построении классификации признаки выражены через абсолютные величины, а более подробную информацию о них можно получить из ряда работ /Караванов, 1980; Sneath, Sokal 1973; и др./.

Для признаков, принимающих значение непрерывной области, наибольшее распространение получила взвешенная Евклидова метрика

$$d_{i,j} = \left| \sum_{k=1}^{m} w^{(k)} \left(x_i^k - x_j^k \right) \right|^{1/k}; \quad \sum_{k=1}^{m} w^{(k)} = 1. \qquad (4.6)$$

Взвешивание исходной системы признаков достигается путем их центрирования и нормирования относительно среднего по совокупности данных и дисперсии. С помощью такой процедуры осуществляется переход признаков к безмерным величинам, что позволяет учитывать каждый признак при классификации *с одинаковым весом.*

Показатель степени k может быть любым числом (большим или равным 2). Чаще всего используют квадратическую форму, но, если различие объектов по какому-то из признаков более существенно, чем сумма различий между всеми остальными признаками, Ю.В. Николаев /1976/ рекомендует показатель степени k брать больше 2. В этой же работе приводятся еще два интересных выражения мер сходства:

$$S(x_i, x_j) = \frac{1}{1 + \alpha d(x_i, x_j)}, \qquad (4.7)$$

$$S(x_i, x_j) = \beta E^{-\alpha d(x_i, x_j)}, \qquad (4.8)$$

где $d(x_i, x_j)$ определяется по приведенным формулам, а параметры α и β выбираются произвольно. На наш взгляд, два последних выражения более гибки, чем метрики Евклида или Махалонобиса. Последняя имеет следующий вид:

$$d_{i,j} = \sqrt{(x_i - x_j) \sum w' (x_i - x_j)}, \qquad (4.9)$$

где ' – символ транспонирования, W – ковариационная матрица. Это выражение желательно использовать при существенных различиях в значимости признаков. Наибольшая часть веса в этом случае приписывается более информативным признакам.

Гибкость применения для классификации выражений (4.7) и (4.8) достигается за счет параметров α и β, которые могут быть подобраны путем использования искусственных эталонов с различными свойствами и структурой, отражающими особенности распределения объектов в признаковом пространстве. Приведенные метрики не исчерпывают имеющийся арсенал расстояний, и их применение должно в первую очередь диктоваться внутренней структурой признакового пространства и целью классификации. С точки зрения внутренней структуры и формальной организации автоматические алгоритмы классификации подразделяются на иерархические и неиерархические /Дюран, Оделл, 1977/.

Иерархические, в свою очередь, подразделяются на агломеративные и дивизимные. В первых алгоритмах начальная процедура объединения состоит из одноэлементных классов, находящихся в непосредственной близости друг от друга. Затем, по мере последующих итераций, объединяются более отдаленные элементы, и в конечном итоге все элементы объединяются в один класс на каком-то уровне классификации. Работа дивизимных алгоритмов проводится в обратной последовательности – от множества к элементам. Достоинство иерархических методов классификации состоит в их наглядности и возможности проводить изучение внутренней структуры множества на различных уровнях объединения. Недостатки сводятся к громоздкости вычислительных операций и необходимости обладания емкой памятью вычислительных машин.

В неиерархических кластер-процедурах в качестве основного принципа образования кластеров также выступает сгущение точек в многомерном пространстве. При этом центры классов должны достаточно далеко удаляться друг от друга, а все точки, объединяемые в классы, находиться в пределах действия какого-либо эталона-моды или медианы. Предполагается, что расстояние между объектом и эталоном одного класса меньше, чем расстояние между этим же объектом и эталоном другого. В качестве решающих правил при выделении кластеров выступают поверхности; чаще всего это куб, плоскость, сфера. Наибольшее распространение эти плоскости получили в виде гиперсферы с заданным определенным радиусом; при работе автоматических алгоритмов классификации они останавливаются в местах сгущения точек и отыскивают границы классов. Такие операции сопровождаются на каждой остановке подсчетом центров тяжести и наиболее интенсивным перемещением в направлении сгущения точек. Перемещение гиперсферы продолжается до тех пор, пока не перестанут изменяться координаты центров тяжести; в этом случае класс считается выделенным и исключается из дальнейшей классификации. В результате таких процедур происходит выделение непересекаемых классов, число которых заранее не задается. Точки, которые не попали в классы, считаются шумом при распознавании образов.

Зависимость структуры выделяемых классов от величины радиуса гиперсферы – недостаток неиерархических алгоритмов. При больших значениях радиуса образуется малое число классов, при малых – наоборот, каждая точка образует самостоятельный класс. Этот недостаток легко устраним либо путем многократной прогонки исходных данных с различными значениями радиуса, либо путем его поиска в смысле оптимальных размеров на искусственных эталонах. Достоинство этих алгоритмов – простота, а специальные исследования показали, что специалист при классификациях пользуется: наиболее простыми решающими правилами; возможностью использования для целей классификации практически неограниченного количества информации по признакам и совокуп-

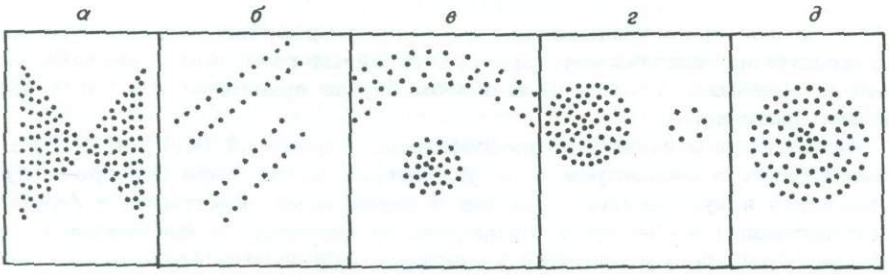


Рис. 14. Примеры распределения точек в признаковом пространстве (по И.Н. Елисеевой и В.О. Руковишникову /1977/) с дополнениями и изменениями.
Кластеры: а – соединяющиеся, б – несферической формы, в – линейно-несипарабельные, г – с большим различием плотности и объемов, д – практически не классифицируемые.

ности объектов; выделением более сложных (по сравнению с иерархическими алгоритмами) классов; непосредственным получением статистических характеристик выделенных классов и даже отображением этих классов на карте (районирование).

При моделировании гидрогеологических систем мы имеем дело с диффузными, плохо организованными системами. Это свойство особенно ярко проявляется при построении автоматических классификаций, когда алгоритмы не дают устойчивых решений. Наиболее распространенные примеры расположения точек в пространстве признаков, при которых возникают неустойчивые решения, приведены на рис. 14.

Распределение точек в пространстве признаков, число классов и их образы, как правило, неизвестны перед началом классификации. Тем не менее определенную информацию о статистической структуре признаков и характере классов можно получить предварительно, с помощью элементарных статистических методов и методов многомерного анализа (например, факторного). Проверка классификационных свойств последнего на искусственных эталонах и реальных гидрогеологических объектах осуществлялась неоднократно. В случае необходимости проведения классификации при неустойчивых решениях алгоритмов можно идти двумя путями.

Первый предусматривает усложнение решающих правил классификации. Один из таких алгоритмов, основанный на представлении размытых множеств, позволяет преодолеть некоторые трудности при классификации сложных по форме кластеров /Елисеева, Руковишников, 1977/. Усложнение решающих правил классификации в основном идет по пути оптимизации процессов поиска точек, объединяемых в признаковом пространстве в один кластер. Это приводит к затратам значительных средств машинного времени и к сложности взаимоотношений между исследователем и ЭВМ.

Второй путь (на наш взгляд, более предпочтительный) сводится к качественному и всестороннему анализу исходных данных. Он должен включать пересмотр, уточнение, проверку исходной информации, выделение информативных признаков, возможное расширение их числа и числа классифицируемых объектов, переход к комплексным более информативным характеристикам или другим метрическим шкалам. Таким путем может быть достигнута наибольшая адекватность классификации природных систем, он не исключает поиск новых решающих правил автоматической классификации и одновременное использование для целей классификации нескольких алгоритмов, использующих различные правила.

Как показывает практика работы одновременно с несколькими алгоритмами, при использовании различных решающих правил на первом этапе выделяются основные кластеры, четкость выделения которых не вызывает сомнения и прослеживается всеми алгоритмами. Последующие просчеты по всей совокупности объектов и отдельно для подмножеств позволяют в значительной мере уточнить границы классов, их число и иерархию. В результате таких расчетов на одних и тех же массивах исходной информации, но разными алгоритмами классификации может создаться ситуация с несколькими вариантами. Выбор приемлемого варианта остается за специалистом, который с помощью классификаций пытается объяснить закономерности интересующего процесса. Обычно выбирается тот вариант, который легко интерпретируется с точки зрения концептуальных моделей или логических построений перед решением задачи. Однако возможны варианты, когда сами итеративные построения классификаций и применение различных алгоритмов открывают новые характерные черты интересующего процесса, а при исследовании природных систем ставят новые задачи.

Конечный эффект работы любого алгоритма автоматической классификации должен оцениваться с точки зрения содержательной интерпретации классов, и всегда, при любых построениях классификаций, надо помнить, что моделирование – не самоцель, а средство познания.

4.2. КЛАССИФИКАЦИЯ УСЛОВИЙ ФОРМИРОВАНИЯ РЕСУРСОВ ПОДЗЕМНЫХ ВОД НА ОСНОВЕ ИЕРАРХИЧЕСКОЙ ТАКСОНОМИИ

Кластерный анализ сравнительно давно применяется в геологических науках. Его широкое применение обусловлено относительной простотой, наглядностью изображения результатов классификации, а также отсутствием заранее заданного числа классов и возможностью использования достаточно больших объемов исходной информации. Довольно подробно применение теории этой разновидности методов многомерного анализа описано в ряде работ /Абрамович и др., 1975; Бугаец, Дуденко, 1976; Девис, 1977; Крамбейн, Грейбилл, 1969; Крамбейн и др., 1973/.

В основу кластерного анализа положен принцип оптимального разбиения заданного множества на ряд непересекающихся подмножеств с помощью критерия оптимальности. Чаще всего используется парагрупповой метод, алгоритм которого представляет итерационную процедуру, число шагов которой $N-1$, где N – число элементов множества X. Вначале каждый элемент выборки представляет отдельную группу, а в качестве матрицы межгрупповых расстояний $S^{o} = \{S^{(o)}_{i,j}\}$, $i, j = 1, 2, \ldots, N$ принимается матрица парных расстояний (метрик) R.

На t-м шаге итерации $t = 1, 2, \ldots, N$ объединяются также группы i_o и j_o, на которых достигается минимум

$$T^{(t)}_{i,j} = \frac{S^{(t)}_{i,i} + S^{(t)}_{j,j} + S^{(t)}_{i,j}}{n^{(t)}_i + n^{(t)}_j}, \qquad (4.10)$$

где $n^{(t)}_i$ – число групп, которое ранее объединено в i-ю группу. Этой группе присваивается номер $\min(i_o, j_o)$, если $i_o < j_o$, то

$$S^{(t+1)}_{i_o, j_o} = S^{(t)}_{i_o, i_o} + S^{(t)}_{j_o, j_o} + S^{(t)}_{i_o, j_o}, \qquad (4.11)$$

$$a\, S^{(t+1)}_{i_o, k} = S^{(t)}_{i_o, k} + S^{(t)}_{j_o, k},$$

при этом для всех $k \neq i_o, j_o$.

После этих операций столбец матрицы $S^{(t)}$ с номером j_o исключается из рассмотрения, и на шаге N-1 все объекты объединяются в класс и строится дендрограмма, которая представляет собой граф, напоминающий дерево и используемый для графического изображения взаимных связей между объектами из заданного множества. Иерархический уровень расположения объектов подчеркивает их взаимное сходство на основе измеряемых признаков.

На первом этапе расчетов в качестве исходной выборки принималась та же совокупность из 80 речных бассейнов с 14 признаками (см. гл. 3), а в качестве метрики – Евклидово расстояние. Результаты расчетов по программе "Корреляция" представлены на рис. 15 в виде дендрограммы.

Средние обобщенные расстояния между объектами изменяются от 0,16 до 0,47; наиболее часто встречаются значения в интервале от 0,17 до 0,22; вероятность попадания обобщенных расстояний в этот интервал – 0,7. Объединение в классы начинается с метрики сходства 0,012 и заканчивается на уровне 0,36. Класс считался выделенным, если число объектов, содержащихся в нем, составляло не менее двух.

Рис. 16 позволяет проследить динамику объединения объектов в классы на различных уровнях (от 0,04 до 0,28) метрик сходства. При минимальной метрике 0,02 в классы объединяются гидрометрические посты, находящиеся на одной реке в непосредственной близости друг от друга. На этом уровне формируются лишь два класса: первый включает водопосты в низовьях р. Хилок (3 объекта), второй охватывает низовья р. Уда (2 объекта), что составляет около 8% от общей совокупности. Для этих районов свойственно формирование подземного стока в условиях широкого развития межгорных артезианских бассейнов. С позиций формирования ресурсов подземных вод по поясам в горно-складчатых областях выделяемые районы характеризуются преимущественно долинными условиями. Для этих районов типично постоянное подземное питание рек в течение года и в многолетнем разрезе. Стабилизации режима ресурсов способствует мощная толща рыхлых четвертичных отложений, активная связь напорных вод (из нижележащих водоносных горизонтов) с грунтовыми, на отдельных участках не исключена возможность берегового регулирования.

Перечисленные районы на этом уровне образуют так называемые ядра классов, к которым по мере увеличения метрики сходства будут присоединяться другие объекты. На определенных этапах дальнейшей классификации эти ядра могут сливаться с другими, образуя новые классы (см. рис. 16).

На уровне меры сходства 0,04 образуется уже семь классов, объединивших 21 объект, что составляет 25% от общего числа. Наряду с бассейновым принципом объединения таких крупных водотоков, как Баргузин, Курба и другие, начинают формироваться приводораздельные межбассейновые районы, охватывающие верховья водотоков со сходными условиями формирования ресурсов подземных вод. Последние районы по характеру формирования подземного стока резко отличаются от первых, и для них типичны условия массивов. Например, на этой стадии объединяются верховья рек в бассейнах Уды и Хилка, где формирование ресурсов подземных вод происходит при отсутствии мощных

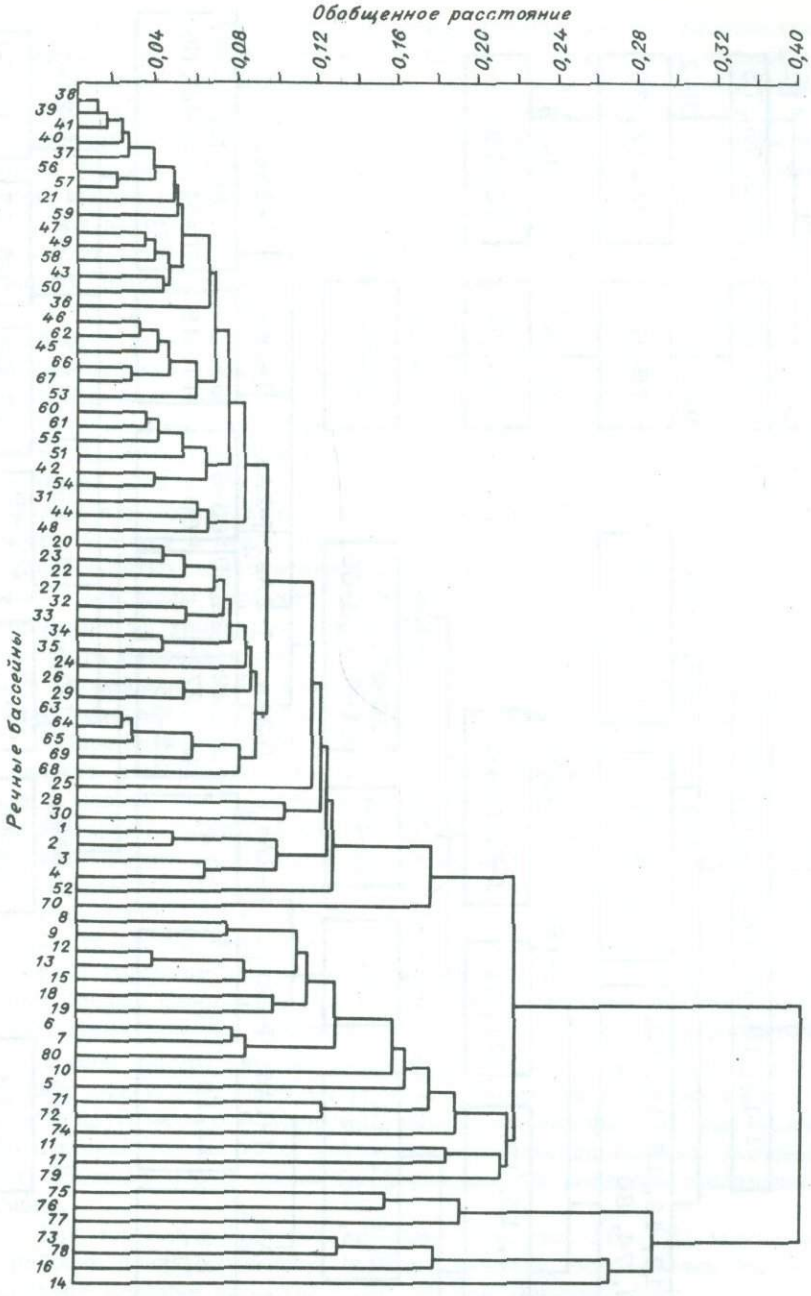


Рис. 15. Дендрограмма кластер-анализа Q -типа.

толщ четвертичных отложений, больших уклонов местности и повсеместного развития многолетнемерзлых пород. Такие условия порождают значительную изменчивость и малые величины основных параметров подземного стока. В течение зимнего периода на отдельных водотоках подземный сток в реках может отсутствовать.

Заметная дифференциация сохраняется на метрике сходства 0,08, при которой в 10 классов объединилось 57 объектов, или 70% общего числа. Наи-

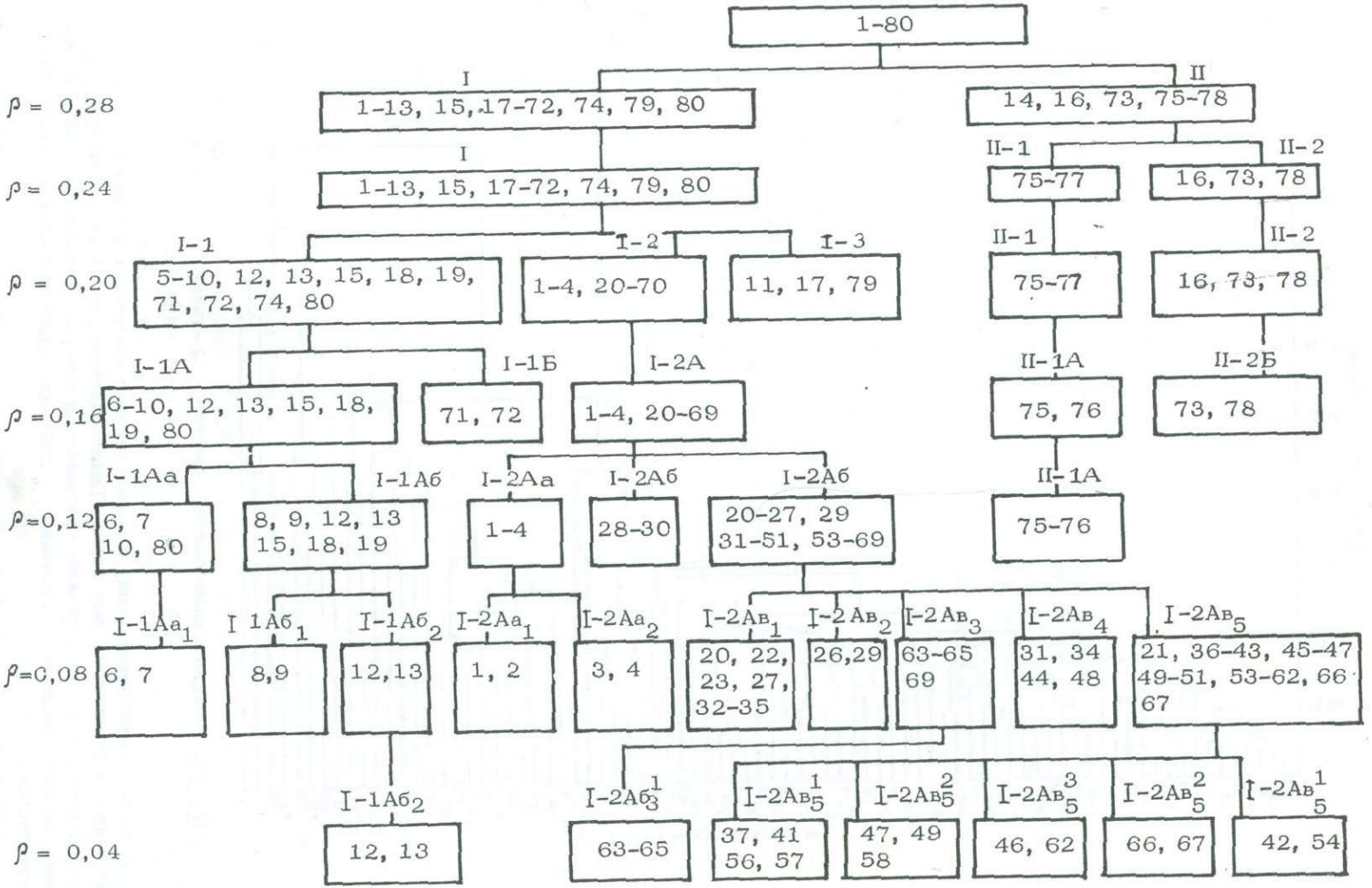


Рис. 16. Структура объединения объектов в классы при различных величинах метрики сходства (ρ).

меньшие по числу объектов классы характерны для рек Байкальской рифтовой зоны. Большинство рек бассейна Селенги на этой стадии сходства объединилось в довольно устойчивые районы, которые на последующей стадии 0,12 объединяются всего в два класса. Высокая дифференцированность для районов Прибайкалья сохраняется вплоть до последней итерации. Сказанное свидетельствует, что изменчивость условий формирования ресурсов подземных вод для этого региона гораздо выше, чем в бассейне р. Селенга.

При переходе к мере сходства 0,16 число объединяемых в классы объектов составляет 90% общего количества. На этой стадии уже виден территориальный принцип объединения более мелких иерархических единиц; особенно хорошо это прослеживается при объединении, соответствующем мере сходства 0,2. Так, в классы I-1, I-3 и II-1, II-2 входят бассейны подземного стока преимущественно Байкальской рифтовой зоны, за исключением участка западного побережья: класс I-2 включает реки бассейна Селенги.

Индивидуальность классов особенно ярко проявляется на примерах рек Южного Прибайкалья, где районы имеют отличные черты формирования подземного стока. Район II-2 - это бассейны рек, где ресурсы подземных вод формируются в условиях максимального увлажнения, а высокая проницаемость атмосферной влаги в подземные воды обусловлена широким развитием интенсивно трещиноватых карбонатных пород. Именно здесь сосредоточены максимальные (по основным параметрам) ресурсы подземного стока. При последующем объединении на уровне метрики сходства 0,24 получается три класса, при этом часть бассейнов рифтовой зоны образует совместный класс с селенгинскими бассейнами, а на конечной стадии объединения 0,28 выделяются два класса, которые при метрике 0,365 объединяются в одну совокупность.

Приведенная итерационная процедура классификации показывает неопределенность в выборе конечного числа классов. Интенсивность объединения объектов в классы при различных метриках сходства на используемом эмпирическом материале показана на рис. 17 и подчиняется экспоненциальному закону. Эта закономерность исследовалась при изменении признакового пространства до четырех-пяти членов. Вторая классификационная процедура (как и последующая) проводилась также на 80 объектах со следующими признаками: модуль подземного стока, величина увлажнения, комплексный показатель расчленения рельефа и градиент вертикальных неотектонических движений земной коры и уклона речных бассейнов. При классификации по четырем признакам градиент вертикальных неотектонических движений земной коры из расчетов исключался.

Выбор такого комплекса признаков обусловлен наличием высоких факторных нагрузок в факторной матрице и сильных связей с модулем подземного стока и другими его основными параметрами-коэффициентами подземного стока и питания. Кроме того, названные признаки определены независимо и отражают основные черты процессов формирования ресурсов подземных вод в регионе.

При классификации совокупности (см. рис. 16, 17) можно принять схему районирования, состоящую из пяти районов, основываясь при этом на минимальной метрике сходства 0,20. Выбор метрики обусловлен расчетами на обучающейся выборке (см. разд. 4.4). На этой стадии полнота охвата классификацией объектов составляет 79 членов совокупности, и лишь один объект (бассейн р. Улюкчикан) единичен. Полученная схема районирования для Байкальской горно-складчатой области отражена на рис. 18.

Следует отметить одну характерную деталь: вся область Селенгино-Витимского прогиба на этом уровне классификации объединена в один класс, хотя на меньших метриках сходства регион дифференцируется на ряд мелких таксономических единиц.

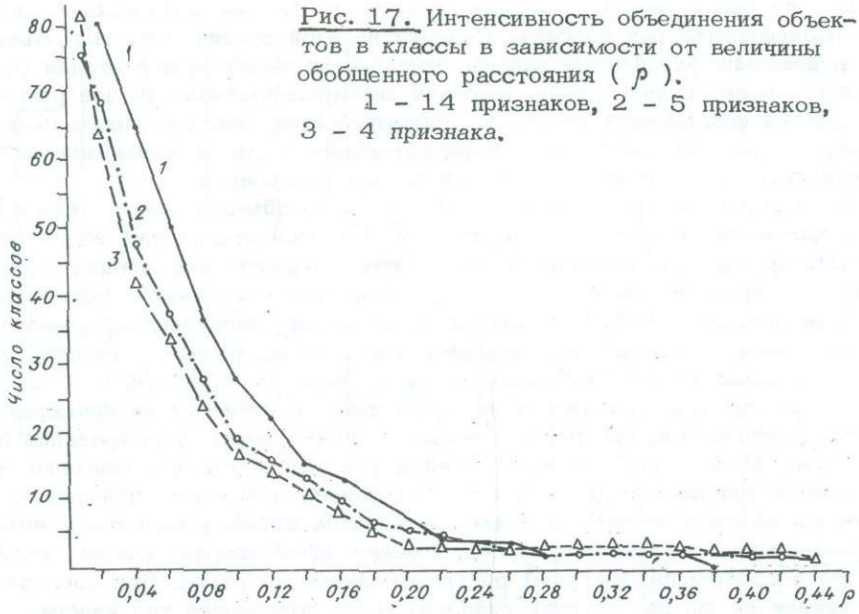


Рис. 17. Интенсивность объединения объектов в классы в зависимости от величины обобщенного расстояния (ρ).
1 - 14 признаков, 2 - 5 признаков, 3 - 4 признака.

До сих пор мы имели дело с Q-схемой кластерного анализа, основная задача которой - типизация в многомерном пространстве объектов. Существует также R-схема, основная цель которой - рассмотрение структуры связей в системе признаков. В качестве меры сходства здесь рассматриваются коэффициенты корреляции между признаками, эта процедура также носит название "анализ групп": вначале рассчитывается матрица парных коэффициентов корреляции; следующий шаг - попарное объединение признаков с высокими коэффициентами корреляции; далее рассчитывается новая матрица сходства, а сгруппированные элементы принимаются в качестве единичных, и рассчитываются новые коэффициенты корреляции. Такая итерация продолжается последовательно до тех пор, пока не будут исчерпаны все признаки. Получаемые в результате расчетов новые значения коэффициентов связи носят название кофенетических коэффициентов корреляции. Существует довольно много методов группировки признаков и расчетов кофенетических коэффициентов сходства /Девис, 1977; Дюран, Оделл, 1977; и др./.

Следует отметить, что построение дендрограмм сходства признаков возможно и на основе меры расстояния, которая может быть рассчитана путем стандартизации исходных данных, и затем дальнейший ее расчет аналогичен вычислению коэффициентов корреляции. В отличие от последнего мера сходства может быть больше единицы, и тем теснее связаны признаки, чем меньше ее величина.

На рис. 19 изображена дендрограмма сходства 14 признаков, используемых для классификации той же исходной совокупности 80 объектов. Величина критического коэффициента корреляции при $n = 14$ для уровня значимости 0,05 составляет 0,50, а для уровня значимости 0,01 это значение будет иметь величину 0,62. В качестве нижнего предела сильных связей будем в дальнейшем использовать значение коэффициента корреляции 0,7, а умеренные связи рассматривать в интервале 0,7-0,4. С этих позиций наиболее связаны в рассматриваемой совокупности признаков основные параметры подземного стока с градиентом вертикальных неотектонических движений земной коры (рис. 21).

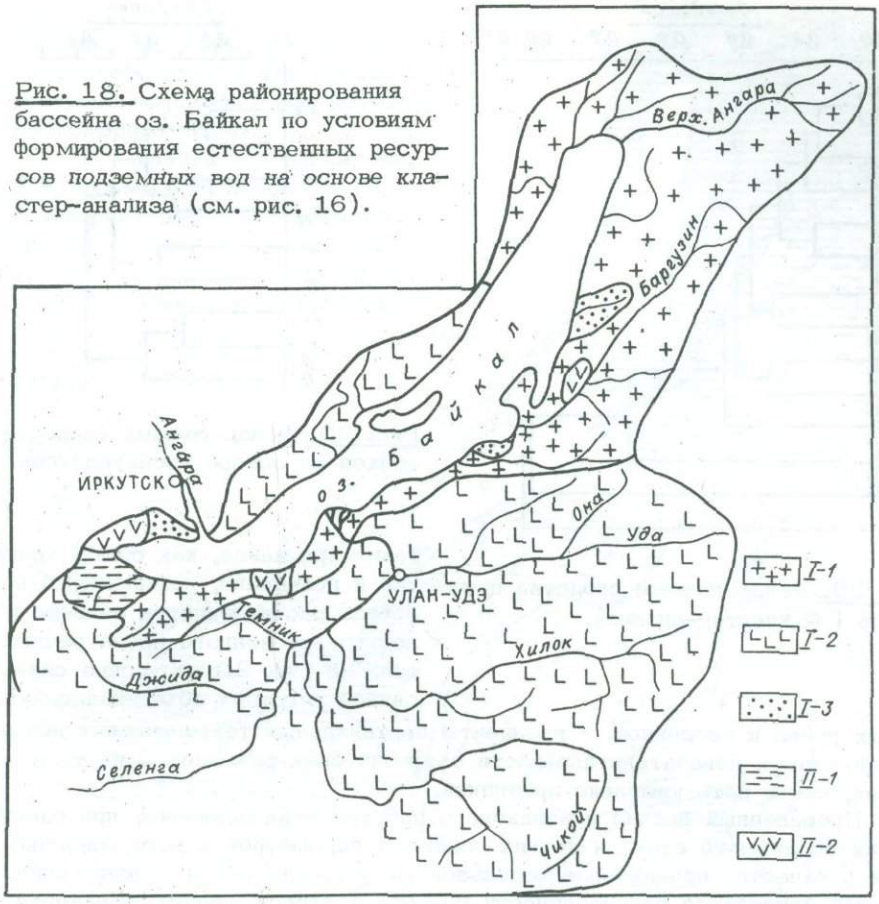


Рис. 18. Схема районирования бассейна оз. Байкал по условиям формирования естественных ресурсов подземных вод на основе кластер-анализа (см. рис. 16).

Самостоятельную группу образуют метеорологические признаки (общее увлажнение и испарение с поверхности водосборов). На уровне сильной связи эта группа объединяется с первой. Величина кофенетического коэффициента корреляции между этими группами 0,71. Между мощностью эрозионного вреза рек и комплексным показателем расчленения рельефа также прослеживается сильная связь. Эти признаки хотя и умеренно, но достаточно надежно (кофенетический коэффициент 0,63) связаны с перечисленными признаками.

В рассматриваемой системе признаков изолированно выступают умеренно связанные с предыдущими группами, такие как плотность тектонических нарушений, средняя высота бассейнов рек, частота речных потоков и уклоны речных русел.

Интересным представляется анализ связей комплекса признаков на основании кластер-процедуры, но с использованием иной метрики сходства. Расчеты проведены для наиболее информативных признаков, при этом параметры подземного стока исключены из расчетов, что дало возможность проследить изолированно характер связей между основными стокообразующими условиями. В качестве меры сходства использовался косинус угла θ (4.16), а результаты расчетов представлены в виде дендрограммы на рис. 20. Сравнительный анализ двух дендрограмм показывает принципиальное сходство в объединении

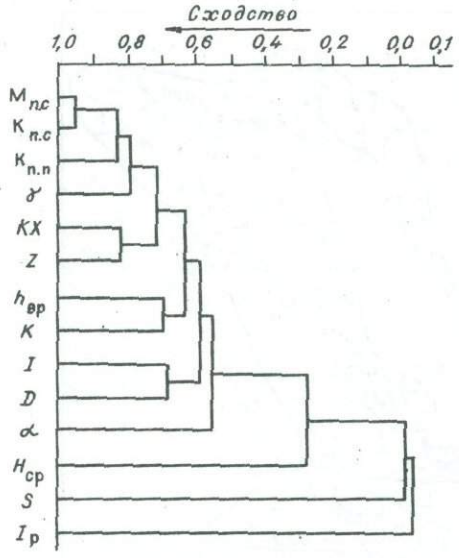


Рис. 19. Дендрограмма сходства признаков (R кластер-анализ).

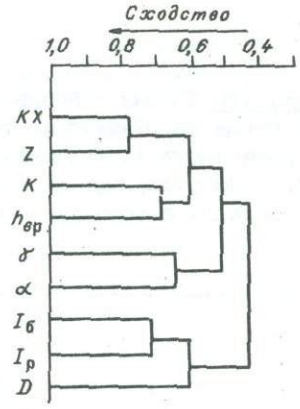


Рис. 20. Дендрограмма сходства признаков на основе косинуса угла.

таких признаков, как общее увлажнение и испарение, комплексный показатель расчлененности рельефа и мощность эрозионного вреза речных русел. На рис. 20 на уровне сильных связей в группу объединились уклоны речных русел и бассейнов, а градиенты вертикальных тектонических движений земной коры и показатели плотности тектонических разломов – на уровне умеренных, как и весь комплекс признаков.

Проведенный анализ признакового пространства позволяет при оценке модуля подземного стока и других основных параметров в регрессионных моделях в качестве предикторов использовать величины общего увлажнения, комплексного показателя расчлененности рельефа и относительного градиента вертикальных тектонических движений земной коры. Именно с помощью этих признаков контролируют пространственную изменчивость основных параметров подземного стока в бассейне оз. Байкал.

4.3. КЛАССИФИКАЦИЯ УСЛОВИЙ ФОРМИРОВАНИЯ РЕСУРСОВ ПОДЗЕМНЫХ ВОД НА ОСНОВЕ НЕИЕРАРХИЧЕСКОЙ ТАКСОНОМИИ

По сравнению с иерархическими методами неиерархическая таксономия в практике геологических исследований начала применяться лишь в последнее время. Трудности использования этой категории методов для решения задач практической классификации, по-видимому, обусловлены сложностью алгоритмов и их практической реализацией на ЭВМ.

В наших примерах автоматическая классификация осуществлялась с использованием той же выборки (что и в кластерном анализе) на ЭВМ ЕС-1022 с помощью алгоритма /Евдокимова, 1978/. Данный алгоритм разработан с целью обработки массивов геолого-геохимической информации и предусматривает реализацию трех программ: "Классификация", "Интерполяция" и "Ассоциация".

Алгоритм "Классификация" строится на основании широко известного метода гиперсфер. В признаковом пространстве выбирается точка Q_o, являющая-

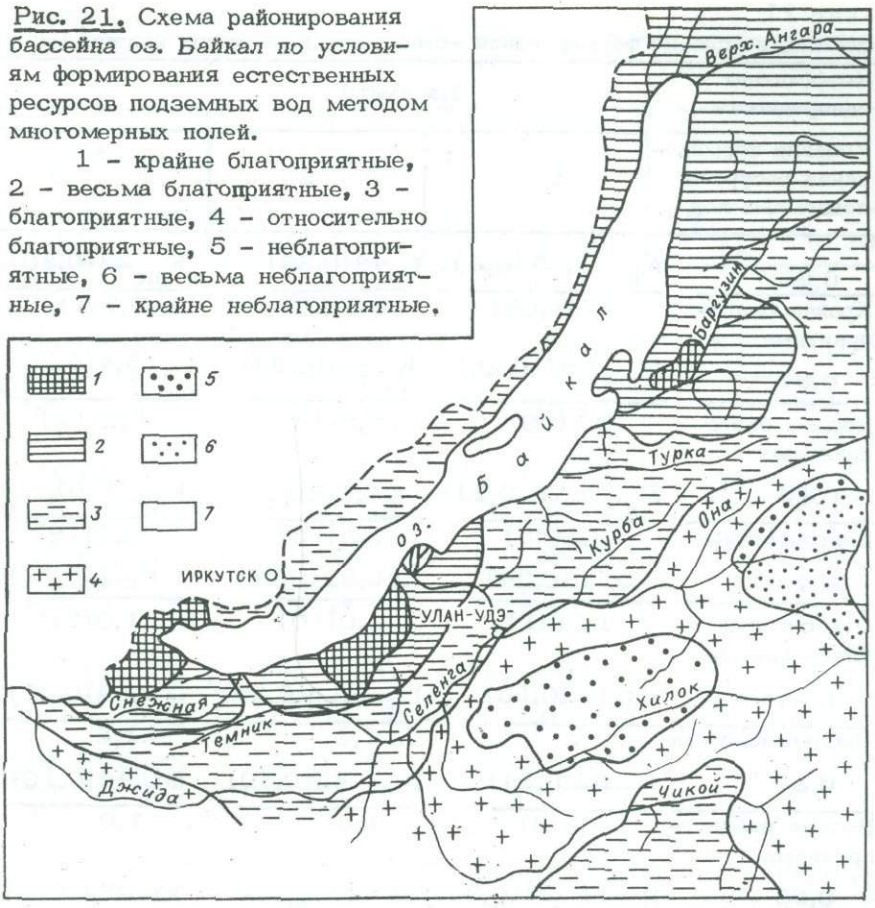


Рис. 21. Схема районирования бассейна оз. Байкал по условиям формирования естественных ресурсов подземных вод методом многомерных полей.

1 – крайне благоприятные, 2 – весьма благоприятные, 3 – благоприятные, 4 – относительно благоприятные, 5 – неблагоприятные, 6 – весьма неблагоприятные, 7 – крайне неблагоприятные.

ся центром начальной гиперсферы, задается радиус R_o и определяются точки-векторы, попавшие внутрь выделенного пространства. Затем находится центр тяжести Q_1 для точек, попавших внутрь этой гиперсферы, при совпадении центров тяжести Q_o и Q_1 группа считается выделенной. Если точки (центры) не совпали, то Q_1 принимается за новый центр гиперсферы и находятся точки, попавшие в гиперсферу радиуса R_o в центром Q_τ. Процедура повторяется до тех пор, пока последующий центр тяжести Q_τ не совпадет с предыдущим, т.е. пока не перестанут изменяться координаты центров тяжести. В этом случае центр гиперсферы будет находиться в области максимума плотности точек, и после объединения в классы они автоматически исключаются из процедуры дальнейшей классификации.

Критерии правильности классификации следующие: точки, принадлежащие к одному классу, должны быть близки между собой в признаковом пространстве; в каждом классе должно существовать одно основное соотношение между признаками (коэффициент однородности); получаемые классы могут делиться до тех пор, пока не будет удовлетворен критерий однородности. Последний основан на следующем условии: остаточная дисперсия признаков, входящих в один класс, не должна превышать некоторую, достаточно малую, предварительно заданную величину. Эта величина выражается через коэффициент однород-

Таблица 11

Классификация условий формирования естественных ресурсов подземных вод

Класс	Коэффициент однородности / Условия формирования ресурсов подземных вод	Параметр 1	Параметр 2	Параметр 3
1	0,99 / Крайне благоприятные	$M_{п.с}$ 8,65 (0,23) / 17,3(63)	$I_{р}$ 48(0,35) / 9,6(16)	$K_{п.с}$ 23(0,30) / 7,7(11)
2	0,91 / Весьма благоприятные	$M_{п.с}$ 4,86(0,26) / 9,7(66)	$K_{п.с}$ 16(0,24) / 5,3(16)	$I_{р}$ 15,4(0,70) / 5,2(12)
3	0,85 / Благоприятные	$M_{п.с}$ 1,70(0,24) / 3,5(45)	γ 0,86(0,27) / 2,8(15)	$K_{п.с}$ 8 (0,20) / 2,5(18)
4	0,73 / Относительно благоприятные	S 0,10(0,34) / 1,2(21)	γ 0,32(0,44) / 1,0(10)	D 0,43(0,18) / 1,0(5)
5	0,87 / Неблагоприятные	S 0,12(0,10) / 1,2	$I_{р}$ 6,16(0,35) / 1,2	D 0,54(0,07) / 1,2
6	0,87 / Весьма неблагоприятные	$I_{р}$ 32(0,54) / 3,3(8)	$I_{б}$ 145(0,30) / 1,0	D 0,45(0,10) / 1,0
7	0,93 / Крайне неблагоприятные	$H_{с.р}$ 1028(0,04) / 1,0	Z 413(0,07) / 0,9	KX 453(0,04) / 0,8

ности (η); чем больше η, тем однороднее класс и меньше вероятность, что он объединяет несколько подклассов. Удовлетворительные решения классификации характеризуются коэффициентами однородности 0,75-0,95, расчет базируется на основании определения собственных значений ковариационной матрицы признаков. Максимальное собственное значение, удовлетворяющее определенным заранее заданным условиям, является критерием выделения классов. Выделенные классы в свою очередь также характеризуются следующими параметрами: коэффициентами контрастности (КК), представляющими отношение абсолютных значений признаков в классах к фоновой величине, для рассматриваемой области; средними значениями признаков; среднеквадратическими отклонениями и коэффициентами вариации.

Для анализа классов на однородность используется метод главных компонент в целом для всей совокупности признаков (значимость параметров) и для каждого класса отдельно с целью определения веса признака или их информационной значимости. Достоинства данного алгоритма заключаются в том, что классификационные правила базируются на изменчивости всей совокупности признакового пространства (см. гл. 2) и возможности выделения комп-

в бассейне оз. Байкал

Параметр			
4	5	6	7
γ6(0,10)	S 0,40(0,63)	K 719(39)	$I_б$391(0,23)
6,0	4	3	2,7
γ1,43(0,39)	KX 974(0,23)	$h_{вр}$ 381(0,46)	K 444(0,54)
4,8	2,1	2,1	2,0
$h_{вр}$ 363(0,31)	K 340(0,37)	$I_р$ 7,1(0,55)	α 23,6(0,21)
1,5	1,4	1,4	1,3
$K_{п.с}$ 4(0,20)	$M_{п.с}$ 0,55(0,20)	$h_{вр}$ 250(0,17)	$H_{с.р}$ 1077(0,12)
1,1	1,1	1,1	1,0
α 20(0,11)	$I_б$ 149(0,18)	KX 544(0,11)	$H_{с.р}$ 1009(0,03)
1,1	1,0	1,0	1,0
$H_{с.р}$ 986 (0,06)	S 0,08(0,40)	Z 423(0,06)	α 16,6(0,29)
1,0	1,0	1,0	1,0
D 0,31(0,16)	$I_б$ 99(0,38)	α 12,0 (0,59)	$I_р$ 293(0,36)
0,7	0,7	0,60	0,60

лекса (ассоциаций) наиболее информативных признаков на основании их веса и коэффициента контрастности в каждом классе, что дает возможность проследить структуру выделяемых классов. Эти классы представлены в виде карт обобщенных районов и карт отдельных признаков. Процедура классификации поясняется таблицей основных параметров классов (табл. 11). Признаки в классах расположены в порядке уменьшения коэффициентов контрастности. Для расчета последних использовались фоновые значения признаков, которые принимались на основе закона их статистического распределения (вида, асимметрии, математического ожидания, вариации).

С учетом специфики формирования естественных ресурсов подземных вод в Байкальской горно-складчатой области, наиболее адекватной представляется схема районирования, состоящая из семи классов – от крайне благоприятных до крайне неблагоприятных (см. табл. 11). Границы первого – третьего районов от крайне благоприятных до благоприятных условий формирования ресурсов подземных вод в геолого-структурном положении совпадают с границами Байкальской рифтовой зоны (рис. 21).

Иерархические ряды признаков (по коэффициентам контрастности) для

Окончание табл. 11

Класс	Коэффициент однородности / Условия формирования ресурсов подземных вод	Параметр 8	Параметр 9	Параметр 10
1	0,99 / Крайне благоприятные	$K_{п.п}$ 52 (0,33) / 2,6	KX 1232(0,22) / 2,3	D0,93(0,40) / 2,1
2	0,91 / Весьма благоприятные	λ 30,9(0,22) / 1,9	$K_{п.п}$ 43(0,44) / 1,8	S 0,19(0,82) / 1,7
3	0,85 / Благоприятные	$K_{п.п}$ 23(0,23) / 1,2	S 0,13(0,35) / 1,3	KX 700(0,15) / 1,2
4	0,73 / Относительно благоприятные	$K_{п.п}$ 20(0,28) / 1,0(5)	Z 455(0,44) / 1,0	K 222(0,18) / 0,9
5	0,87 / Неблагоприятные	Z436(0,01) / 1,0	K 180(0,24) / 0,8	$h_{вр}$ 165(0,17) / 0,7
6	0,87 / Весьма неблагоприятные	KX 438(0,18) / 0,8	$K_{п.п}$ 15(0,40) / 0,8	$h_{вр}$ 176(0,20) / 0,8
7	0,93 / Крайне неблагоприятные	$h_{вр}$ 137(0,10) / 0,60	γ 0,17(0,23) / 0,60	S 0,04(0,29) / 0,40

Примечание. В числителе – символ признака, его средняя величина, в скобках – коэффициент вариации признака в классе; в знаменателе – коэффициент контрастности, в скобках – значимость признака или вклад в общую дисперсию в пределах выделенного класса.

этих классов позволяют выявить следующие закономерности формирования подземного стока в регионе. Величина увлажнения в первом – третьем районе значительно превышает фоновые величины горно-складчатой системы оз. Байкал (КК = 3,0-1,2). Годовые суммы атмосферных осадков высокогорных районов достигают более 1400 мм. Экстремальных значений в этих районах достигают такие признаки, как градиент скорости вертикальных движений земной ко-

Параметр			
11	12	13	14
$h_{вр}$ 387(0,09) / 1,4	α 25(0,04) / 1,4	Z 597(0,14) / 1,3	$H_{с.р}$ 1247(0,13) / 1,2
$I_{б}$ 207(0,41) / 1,6	D 0,59(0,32) / 1,4	Z 523(0,13) / 1,2	$H_{с.р}$ 1113(0,22) / 1,2
$H_{с.р}$ 1261(0,14) / 1,2	$I_{б}$ 164(0,37) / 1,1	KX 470(0,15) / 1,0	D 0,44(0,22) / 1,0
$I_{б}$ 138(0,21) / 0,9	α 15,2(0,28) / 0,8	KX 508(0,11) / 0,8	$I_{р}$ 2,9(0,51) / 0,6(49)
$M_{п.с}$ 0,35(0,18) / 0,7(10)	$K_{п.п}$ 14(0,17) / 0,7(10)	$K_{п.с}$ 2 (0,0) / 0,7(10)	γ 0,14(0,0) / 0,5(51)
γ 0,21(0,55) / 0,8	K 150(0,15) / 0,6(8)	$K_{п.с}$ 1(0,33) / 0,4(42)	$M_{п.с}$ 0,20(0,41) / 0,4(29)
K 85(0,26) / 0,40(11)	$K_{п.с}$ 1(0,0) / 0,30(10)	$K_{п.п}$ 6(0,20) / 0,30(14)	$M_{п.с}$ 0,10(0,16) / 0,20(46)

ры (КК = 2,8–6,0) и плотность тектонических нарушений (КК = 1,3–1,9), а также комплекс характеристик расчлененности рельефа и уклоны русел рек и речных водосборов.

Общей закономерностью этих районов является объединение в таксономические подразделения бассейнов рек, имеющих совокупность признаков, которая значительно превосходит по своим величинам их фоновые величины для всего региона.

Особенно следует остановиться на процессе испарения, так как его абсолютные значения для первых трех классов, казалось бы, максимальны, что должно найти свое отражение на основных параметрах естественных ресурсов подземных вод. Однако, *анализируя данные табл.* 11, *можно сделать следующий* вывод: увлажненность Байкальской рифтовой зоны настолько велика, что испарение существенно сказывается лишь на формировании поверхностного стока. На величину же подземного стока и его пространственное перераспределе-

ние в территориальных границах трех классов этот процесс заметного влияния не оказывает, поскольку скорости инфильтрации атмосферных осадков в подземные воды достаточно велики в силу высокой степени тектонической раздробленности горных пород.

Указанная схема формирования естественных ресурсов подземных вод Байкальского рифта подтверждается высокими количественными показателями: средняя величина модуля подземного стока для выделенных районов изменяется от 8,65 до 1,70 л/(с·км^2) при КК, достигающих максимальных значений 17,3 (крайне благоприятные условия), коэффициенты подземного стока изменяются от 8 до 23% и имеют значения 23-52% при КК = 2,6-1,2.

К благоприятным, по условиям формирования естественных ресурсов подземных вод, относится также Чикойская возвышенность (северная часть Хэнтэй-Даурского неотектонического поднятия). Условия формирования подземных вод здесь в целом схожи с таковыми Байкальской рифтовой зоны, но меньшее количество выпадающих атмосферных осадков предопределяет более низкие показатели подземного стока.

Принимая во внимание значимость признаков по коэффициентам контрастности (*иерархический ряд признаков внутри класса*), можно отметить, что для первого класса главенствующими являются модули подземного стока, уклоны русел рек и производная характеристика от модуля и величины общего увлажнения – коэффициенты подземного стока. Суммарная дисперсия этих признаков для первого класса около 90%. Остальная часть дисперсии равномерно распределилась на последующие показатели (принцип компонентного анализа). Для второго класса ведущими являются те же самые признаки с суммарной дисперсией 94% при возрастающей роли коэффициента подземного стока. Доминирующими признаками третьего класса являются (как и предыдущего) модули и коэффициенты подземного стока, а место уклонов русел рек занимает относительный показатель скорости тектонических движений земной коры. Суммарная дисперсия перечисленных признаков в этом классе 78%.

Переходя к формированию естественных ресурсов подземных вод в неблагоприятных условиях, отметим, что выделенные районы соответствуют бассейну р. Селенга. К этому же региону принадлежит территория с относительно благоприятными условиями, которая является переходной от благоприятных к неблагоприятным. Значения коэффициентов контрастности для четвертого района по своей величине близки к 1, т.е. к фоновым значениям всего региона Байкальской горно-складчатой области. По значимости признаков в этом классе главенствующая роль отводится морфометрическим характеристикам.

Пятый – седьмой районы характеризуются условиями формирования ресурсов подземных вод от неблагоприятных до крайне неблагоприятных. В целом для этих классов КК признаков близки к 1 или значительно ниже (*особенно для крайне неблагоприятных условий*).

Среднемноголетняя величина увлажнения этих районов изменяется в пределах 550-450 мм при возрастающей роли испарения как с земной поверхности, *так и из зоны аэрации. Испарение может достигать* 440 мм, и, следовательно, лишь небольшая часть инфильтрующихся атмосферных осадков достигает грунтовых вод. К этому следует добавить снижение (по сравнению с рифтовой зоной) тектонической активности территории и региональное ухудшение фильтрационных свойств осадочных отложений мезозойских межгорных впадин, которые здесь занимают значительные пространства. Ресурсы подземных вод, формирующиеся в закрытых и полузакрытых впадинах, имеют слабую гидравлическую связь с реками. Меньшие уклоны местности, более слабая расчлененность рельефа (по сравнению с Байкальской рифтовой зоной) обусловливают низкую динамичность подземного стока и его незначительную роль в общем

речном стоке. Наиболее контрастно такие условия проявляются в седьмом районе.

Несовпадение границ выделенных районов с границами речных водосборов на рис. 21 вызвано особенностью применяемого алгоритма аппроксимации многомерного поля на ЭВМ. Кроме того, для построения этой карты задавались координаты речных водомерных постов, для которых проводился расчет комплекса признаков.

Сопоставление выделенных районов с помощью данного алгоритма и анализ осредненных признаков внутри классов в целом показали достаточно хорошую сходимость с основными положениями формирования естественных ресурсов подземных вод Байкальской горно-складчатой области /Естественные ресурсы..., 1976/.

4.4. ОЦЕНКА ЭФФЕКТИВНОСТИ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ

Существуют десятки различных методов классификации многомерных наблюдений. Все они в той или иной степени обладают достоинствами и недостатками. Для реальных природных объектов в условиях недостатка информации довольно трудно найти объективные критерии, с помощью которых можно произвести оценку работы классифицирующих алгоритмов. Это прежде всего связано с тем, что в естественных таксономических построениях заранее неизвестна классификационная структура объектов или процессов.

При анализе распределения объектов в признаковом пространстве наблюдается некоторый их разброс вдоль признаковых осей, и лишь в редких случаях мы имеем компактные объединения по классам. Степень компактности выделяемых классов во многом условна, и, как справедливо отмечено в работе Р. Сокала /1980/, общепринятого определения кластеров не существует. Обычно большинство исследователей интуитивно понимает, что элементы одного кластера ближе друг к другу, чем к другим, само объединение объектов в классы подгоняет данные под заранее заготовленную модель.

Поскольку любой исследователь природных процессов стремится создать классификацию, близкую к естественной, а ее создание требует большого числа признаков, перспективными представляются сжатие информации и дальнейшие классификационные построения на основе шкалирования. Здесь мы вплотную подошли к оценкам эффективности работы методов факторного анализа для целей классификации, и с этих позиций весьма актуальны результаты расчетов, проведенные Г.П. Гапонцевым /1980/ и С.М. Чесаловым /1980/.

В этих работах в качестве классифицирующих алгоритмов использовались метод главных компонент и факторный анализ. Исходным материалом (обучающей выборкой) служили искусственные эталоны с заранее заданными статистическими свойствами, приближающимися к реальным геологическим ситуациям. Эти эталоны, разработанные А.И. Гавришиным /1980б/, позволяют решать целый комплекс классификационных и других задач.

Качество построения классификаций А.И. Гавришин предлагает оценивать с помощью ошибок первого (α) и второго (β) рода, а также ошибки отказа (γ):

$$\alpha = \frac{K}{N_a}; \quad \beta = \frac{l}{N_a}; \quad \gamma = \frac{P}{N},$$

где K – число наблюдений, принадлежащих к данному однородному классу a, но ошибочно отнесенных к другим классам; l – число наблюдений, не принад-

лежащих к данному классу а, но ошибочно отнесенных к этому классу; N_a – число наблюдений в однородном классе а по описанию эталона; Р – количество наблюдений, не вошедших ни в один из классов (отказы).

На основании оценок указанных ошибок предлагается следующий метод классификации: высокоэффективные ошибки меньше 5%; эффективные – 5-15; среднеэффективные – 15-25; низкоэффективные – 25-35 и неэффективные – больше 35%. /Гавришин, 1980б/.

Метод главных компонент /Гапонцев, 1980/ показал для эталона ДИС-3-1 четкую локализацию двух группировок А и Б и по ошибкам может быть отнесен к разряду эффективных (ошибки первого рода составили для этих группировок 5-7%, второго – близки нулю, третьего – 6%). При выделении подклассов указанные ошибки близки к приведенным, но отмечается возрастание ошибок первого рода для подкласса $Б_2$.

В статье С.М. Чесалова /1980/ показано, как для эталона ДИС-3-1 в результате единичного расчета с использованием факторного анализа произошло разделение на четыре подкласса. Ошибки первого рода составили от 0 до 5%, второго для всех подмножеств равны нулю, четыре точки практически не были отнесены ни к одному из классов.

Отнесение точек к подмножествам в первом и во втором случае проводилось визуально. Вклады главных компонент в общую дисперсию при анализе эталона ДИС-3-1 распределились следующим образом: на первую – 33,9%, на вторую – 27,5, на третью – 11,6%. Информативность признаков при таком варианте разложения эталона также соответствовала заданной структуре.

Наши расчеты по алгоритму /Двинских, 1975/ показали аналогичные результаты. Эталон ДИС-3-1 состоит из двух классов, каждый из которых делится на два одинаковых подкласса. Эти признаки распределены равномерно, и наблюдения независимы. Весь эталон состоит из 100 объектов, содержащих по 10 признаков. В класс А входят объекты: 1-20 и 51-70, в класс В – 21-50 и 71-100. Подкласс $А_1$ составляют объекты 1-20, а подкласс $А_2$ – 51-70. Подкласс $Б_1$ состоит из объектов 21-50, подкласс $Б_2$ – 71-100.

Как известно, Q-схема факторного анализа предполагает транспонирование матрицы исходных данных, а расчет коэффициентов корреляции проводится между объектами. Здесь сразу же остро встает вопрос о достаточности количества признаков для расчета достоверных значений коэффициентов корреляции. Чаще всего при работе с природными объектами ответ на этот вопрос отрицательный. Перспектива же мгновенного получения результатов классификации без анализа исходной системы признаков на их информативность и достоверность весьма заманчива и зачастую довлеет над исследователем.

В рассмотренном примере (см. гл. 3) подавляющее большинство величин факторных нагрузок на первый фактор превысило 0,90. Для второго фактора эти значения близки к нулю. Сложившаяся ситуация привела к некоторой неопределенности и несоответствию результатов классификации объектов при сравнении с другими схемами. Поскольку на второй фактор получены низкие нагрузки объектов, можно прийти к выводу о существовании единого надкласса, при этом выделение подклассов становится бессмысленным. Полученные сомнительные результаты классификации реальных природных объектов заставили обратиться к искусственным эталонам. Результаты построений коэффициентов разложения в плоскостях двух первых факторов эталона ДИС-3-1 показали аналогичную форму распределения поля точек в виде полукруга и подтвердили предположение о двумодальной форме такого соответствия точек. Оценка качества классифицирующей работы этой разновидности факторного анализа показала низкую его эффективность.

Ошибки первого рода для подклассов $А_1$ и $А_2$ соответственно 50 и 25%, для подклассов $Б_1$ и $Б_2$ - 43%; ошибки второго рода для подклассов $А_1$ и $А_2$ имеют величины 20 и 90%, а для подклассов $Б_1$ и $Б_2$ - 56 и 10%. Для класса А общая ошибка первого рода 38, для класса Б - 43%. Общее число точек, не отнесенных ни к одному из классов, составляет 12%. Следуя градации А.И. Гавришина /1980б/, можно сделать вывод, что данный метод неэффективен и, следовательно, Q-схема непригодна в таком варианте к выявлению таксономической структуры классифицируемых объектов. Хотя данный метод и дает некоторую дополнительную информацию о характере выборочной совокупности, его следует применять в сочетании с неформальными суждениями и в виде непараметрической формы.

Более перспективна для целей классификации R-модификация факторного анализа, но при ее применении необходимо выдерживать ряд положений, основные из которых сводятся к следующим.

1. Перед началом классификации необходим тщательный анализ информации на ее достоверность. Особое внимание следует уделить наиболее вариабельным признакам, а также резко отличающимся значениям каждого признака от средних величин.

2. На основании многократных расчетов выявляется комплекс информативных признаков и удаляются признаки, имеющие информационный шум. При этом необходимо учитывать, что при создании политомических классификаций часть признаков, отнесенных к неинформативным, может превратиться в информативные.

3. В совокупности признаков, используемых для целей классификации, должны присутствовать независимые признаки со своеобразной таксономической структурой. Однако следует учитывать, что большое количество таких признаков может привести к невозможности разделения исходной совокупности на группировки.

4. Используемые алгоритмы факторного анализа должны предусматривать нормирование исходной системы признаков, так как значения факторных нагрузок при определении их информационного веса во многом зависят от единиц и шкал измерения признаков.

5. Вклады в общую дисперсию первых факторов, по которым проводится классификация объектов в многомерном пространстве, должны быть по возможности максимальными. Чем выше эти значения, тем качественнее объекты обособляются в группы.

6. При условии придания группам точек в многомерном пространстве системы понятий (содержания) последним придается смысл класса (района).

В процедуре кластер-анализа не учитывается ряд положений, указанных выше, и в этом ее основной недостаток, но, как указывалось выше, метод обладает и рядом достоинств, которые отсутствуют в факторном анализе. Разумный путь при построении многомерных классификаций состоит в совместном использовании методов.

К недостаткам метода кластер-анализа следует также отнести его неопределенность в выборе конечного числа классов, а также оценку их однородности. На первом этапе анализа дендрограммы визуально можно выявить иерархическую структуру устойчивых классов и в дальнейшем проводить анализ формирования этих классов при различных величинах сходства. В приложении к природным объектам эта процедура по своей сути соответствует их делению (например, провинция - область - округ - район - подрайон и т.д.). Поэтому весьма заманчива оценка эффективности работы этого метода.

Результаты классификации эталона ДИС-3-1 с помощью кластер-анали-

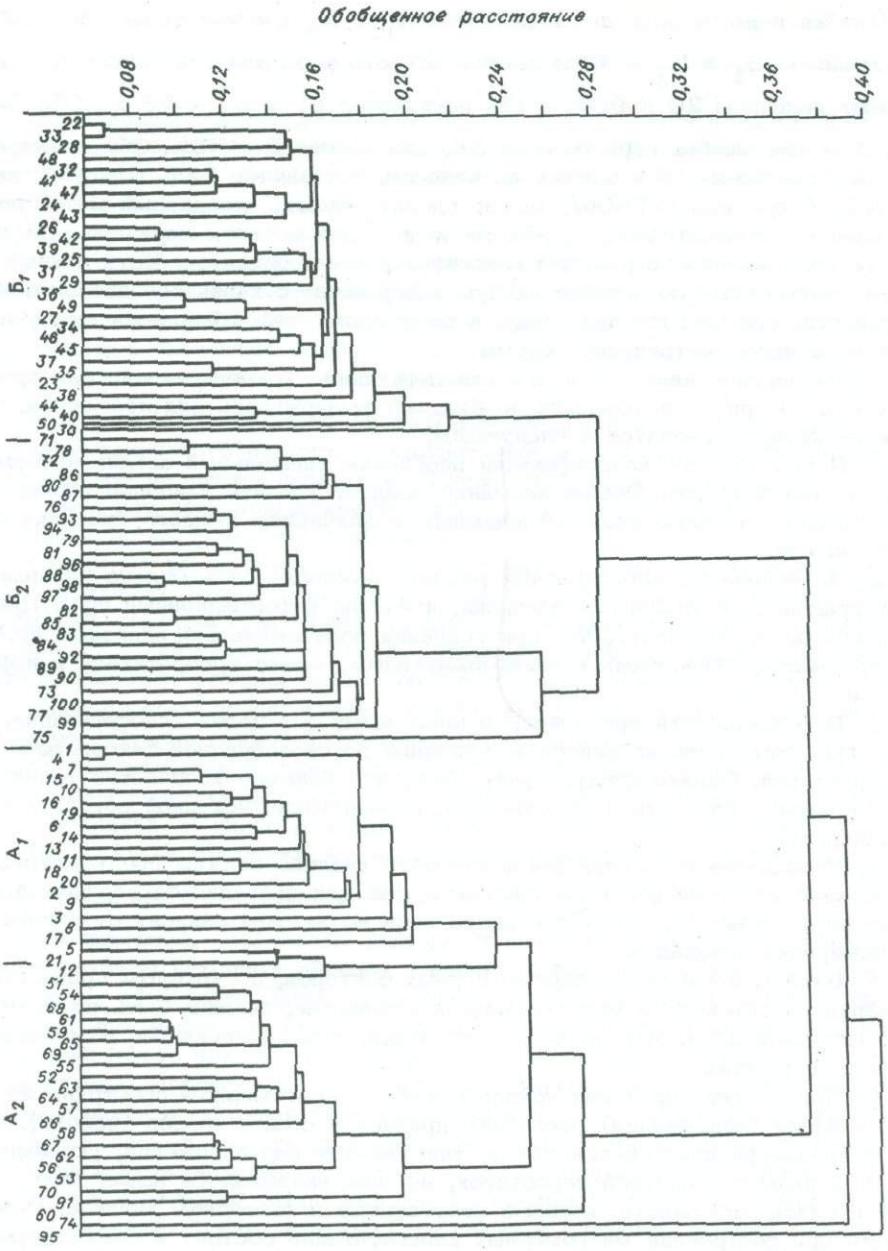


Рис. 22. Дендрограмма кластер-анализа Q-типа для искусственного эталона ДИС-3-1.

за приведены на рис. 22. Анализ дендрограммы показал довольно высокую эффективность применяемого алгоритма. Из подкласса $Б_1$ в подкласс $А_1$ перешел лишь один объект № 21, из $Б_2$ в подкласс $А_2$ - № 91, а объекты № 74, 95 вообще не участвуют в классификации. Общей закономерностью для это-

го эталона является то, что большинство объектов объединяются в классы и подклассы начиная с метрики сходства 0,20 м, заканчивается такое объединение на уровне метрики сходства 0,28. На данной стадии классификации полнота охвата объектов максимальна (98%).

Ошибка первого рода $А_1$ для подклассов $А_1$ и $А_2$ равна 0, для подклассов $Б_1$ и $Б_2$ - 3%; ошибка второго рода для подклассов $А_1$ и $А_2$ - 5%, а для подклассов $Б_1$ и $Б_2$ равна 0. На уровне классов эти ошибки соответственно равны для А - 0, для В - 3%.

Проделанные расчеты позволяют считать применение кластер-анализа для целей многомерных классификаций высокоэффективным.

Использование искусственных эталонов - не единственный методический прием определения конечного числа классов при классификации. Например, /Harfe, Sponholz, 1975/ при исследовании водно-физических свойств грунтов района г. Ростока применялась разработка "Голотип" Ю.А. Воронина на базе обучающих выборок. Этим удалось достичь определенности в формировании классов и отнести сомнительные объекты к тому или иному классу. Разработка "Голотип" в этой работе применялась в комплексе с кластерным анализом.

Наряду с указанными методами для целей классификации многомерных естественных объектов применялся пакет прикладных программ распознавания образов КВАЗАР, разработанный в Институте математики и механики УНЦ АН СССР В.Д. Мазуровым и др. /1978/.

Решение таксономических задач в этом пакете осуществляется "без учителя" с применением трех алгоритмов, использующих модули TAXE, TACLIN и SPECTR. Первый алгоритм таксономии (типа корреляционных плеяд) использует в качестве меры расстояния между классифицируемыми векторами метрику Хемминга. Второй производит разбивку множества на основании параметра R, представляющего значение длины стороны N-мерного гиперкуба (R-объем), используемого для выявления близкорасположенных объектов в пространстве признаков. Таксон при заданном R формируется как совокупность номеров объектов, находящихся внутри объема, образованного пересекающимися R-объемами, и задается в относительных единицах. Последний алгоритм основан на решении потенциальных функций с использованием метрики Евклида с заданным числом классов.

Эффективность работы этих алгоритмов предварительно оценивалась также с помощью искусственных эталонов ДИС-2-1 и ДИС-3-1. Первый эталон аналогичен второму, отличие состоит лишь в том, что признаки распределены по нормальному закону и усилен информационный вес 5-го и 7-го признаков.

Обработка эталона ДИС-2-1 показала, что для алгоритма TAXE классификация основного количества наблюдений на два класса производится при радиусе таксономии R_o, равном 0,48. При этом 14 объектов из 100 образовали единичные таксоны. Рекомендуемый радиус таксономии для этого алгоритма может быть от 0,25 до 0,30. Для программы TACLIN образование двух классов происходит при R_o = 0,30 (7 объектов образовали единичные таксоны).

Ошибки первого рода (α) при использовании алгоритма TAXE с радиусом таксономии 0,479 в этом эталоне для подклассов $А_1$ и $А_2$ составили 5 и 25%; для подклассов $Б_1$ и $Б_2$ - 3 и 15%, что позволяет его считать среднеэффективным, так как общая ошибка выделения класса А - 20%, а для класса Б - 10%.

Алгоритм TACLIN также показал достаточно высокую эффективность.

Ошибки первого рода составили 7,5 и 6,5% для классов, а для подклассов – 1-2%. Ошибки второго рода (β) для подклассов изменяются от 5 до 10%.

Наилучшую (с точки зрения эффективности) классификацию из перечисленных алгоритмов дает программа SPECTR с заданной разбивкой этого эталона на два класса. Величины погрешности первого рода (α) оцениваются в 5%, второго рода (β) – в 10%, при этом показатель качества таксономии составил 3,99.

Расчеты по эталону ДИС-3-1 с помощью двух первых программ для различных радиусов таксономии не дали сколько-нибудь положительных результатов. Программа SPECTR с разделением на четыре подкласса дала погрешности – α = 10%, β = 10% (I подкласс); α = 10%, β = 50% (II подкласс); α = = β = 23% (III подкласс) и α = 27%, β = 20% (IV подкласс).

Объединяя эти четыре подкласса в два класса, получаем α = 2,5%, β = = 7,5% (для I класса) и α = 0,75%, β = 1,7% (для II класса). С заданной разбивкой на два класса всей совокупности эталона ДИС-3-1 ошибка первого рода α = β = 7,5% для I класса и α = β = 5% – для II класса.

Проведенные оценки из группы классификационных алгоритмов пакета показали, что наиболее эффективен для решения таксономических задач алгоритм SPECTR.

После оценки эффективности работы таксономических программ в пакете КВАЗАР произведен выбор оптимальных размеров радиуса таксономии, который в дальнейшем использовался для расчетов на естественных объектах той же совокупности из 80 бассейнов и 14 признаков. Следует отметить, что такой выбор можно производить в случае близкого соответствия данных нормальному закону их распределения.

С помощью алгоритма TAXE при радиусе таксономии 0,46 (близок к выбранному на эталонах) вся совокупность объектов распадается на пять классов (районов) и 16 единичных объектов. В территориальном отношении первый район охватывает бассейн рек Южного и Юго-Западного Прибайкалья, причем из рек южного побережья этот район представляет лишь р. Слюдянка, а все остальные представлены индивидуальными таксонами. Второй район состоит из рек западного побережья среднего Байкала, третий приурочен к северо-западному и северному побережью озера. Четвертый охватывает северо-восточную и восточную (до устья р. Селенги) части бассейна. Аномальными бассейнами выступают реки Рель, Томпуда, Давше, Улюкчикан, Нестериха, а также все реки западного склона хр. Хамар-Дабан. Бассейн р. Селенги образовал пятый самостоятельный район с единичным таксоном (р. Большая – Покровское).

По своей таксономической структуре данная схема близка к схемам районирования, полученным с помощью кластер-анализа, и первому варианту расчетов для всей совокупности объектов на основе R-схемы факторного анализа. В последней схеме подмножества A_2 и A_3 можно рассматривать как состоящие из единичных таксонов. Отдельные аномалии, сформированные в классы, при использовании метода многомерных полей также территориально соответствуют большинству единичных таксонов в пределах Байкальской рифтовой зоны, полученных с помощью алгоритма TAXE.

При использовании алгоритма TACLIN с радиусом таксономии 0,3 вся совокупность бассейнов образует три класса. Первый образовало подавляющее большинство объектов, что территориально соответствует всей площади бассейна оз. Байкал, без разделения на рифтовую зону и Селенгино-Витимский прогиб. Самостоятельные классы образовали реки южного побережья – Утулик и Безымянная, а также Нестериха и Улюкчикан. В этот же класс вошли реки Слюдянка и Аса, хотя территориально они разобщены.

Следует отметить, что данный алгоритм показал наихудшее таксономическое соответствие не только на искусственных эталонах, но и на естественных объектах.

Более интересные результаты, с нашей точки зрения, дал алгоритм SPECTR с заданной разбивкой на семь классов. Число классов выбрано по аналогии с результатами обработки исходной информации методом многомерных полей.

Первый и второй районы территориально совпали с районами, выделенными методами кластер-анализа (номера I-2, I-1), R-факторного анализа ($Б_1$, $А_1$), многомерных полей (3,2) и алгоритма TAXE (I, II, III районы). Такое совпадение результатов выявления этих районов с помощью пяти различных алгоритмов указывает на объективное существование данных таксономических единиц. Как и в кластер-анализе, к первому классу в алгоритме SPECTR отнесен бассейн р. Селенга, но в отличие от первого метода мелкие водосборы этого бассейна (Чикой - Верхний Чикой, Ямаровка, Долентуй) образовали единичные таксоны.

Всеми алгоритмами, кроме Q-схемы факторного анализа, обособляются в один район с аномальными свойствами водосборы рек Нестериха и Улюкчикан (бассейн р. Баргузин): факторный анализ - $А_3$, кластер-анализ - 1-3, метод многомерных полей - 1 (см. рис. 21).

Наибольшее число районов, различающихся по условиям формирования естественных ресурсов подземных вод, характерно для южного (от истока Ангары и до устья Селенги) побережья оз. Байкал. Это нашло свое подтверждение на всех без исключения схемах районирования. Различными алгоритмами здесь выделяется от двух до девяти районов (если принять за район единичные объекты). Анализ закономерностей формирования ресурсов подземных вод подтверждает правомерность дробной разбивки этого региона. Особенно близкие результаты для Южного Прибайкалья дали алгоритмы кластер-анализа, Q-схемы факторного анализа и пакета КВАЗАР.

Если обратиться к схеме районирования горно-складчатых областей по условиям формирования подземного стока (см. табл. 1) и схемам районирования бассейна оз. Байкал, полученным с помощью методов многомерного анализа, можно увидеть, что на уровне таксономического подразделения области почти все используемые алгоритмы четко выделяют Байкальскую рифтовую зону, Селенгино-Витимский прогиб и Хэнтэй-Даурское неотектоническое поднятие. Последняя область не выделяется алгоритмами TAXE, TACLIN и кластер-анализом при метрике сходства 0,20. На более низких значениях метрики сходства возможность выделения этой области очевидна (см. рис. 17). Реальное существование этого таксономического подразделения подтверждают карты факторных поверхностей (см. рис. 10), результаты расчетов с помощью других алгоритмов многомерного анализа и фактические данные по основным параметрам подземного стока (см. табл. 2).

Несколько труднее выделение на предлагаемых схемах округов подземного стока. Поскольку эта таксономическая единица является иерархически соподчиненной областям, то ее адекватность лучше всего сопоставима на кластер-модели с метриками сходства менее 0,20. При анализе рис. 17 можно выявить округа подземного стока не только в таких структурных подразделениях, как горные хребты в зонах избыточного и достаточного увлажеенения, но и горные хребты в условиях недостаточного увлажнения. Сочетания горных хребтов с межгорными впадинами для перечисленных условий увлажнения также довольно отчетливо выделяются на иерархической схеме таксономии (см. рис. 15, 17).

С методологической позиции оперирование таксономическими методами на этом иерархическом уровне с предлагаемым комплексом признаков может

дать наибольший эффект и практически неограниченные возможности методов для решения различных гидрогеологических задач (прогноз ресурсов по комплексу косвенных признаков и др.). Предлагаемая схема районирования горно-складчатых областей по условиям формирования подземного стока (см. табл. 1) не догма и может на данном таксономическом уровне значительно дополняться вследствие выявления новых характерных особенностей формирования естественных ресурсов подземных вод в округах.

При переходе к еще более низким таксономическим подразделениям (районам) использование методов многомерного анализа не менее эффективно, чем для округов. Почти всеми алгоритмами выделяются бассейны рек с развитием карбонатных и интенсивно-трещиноватых алюмосиликатных пород и с другими, отличными от общих региональных закономерностей, свойствами.

Последующие таксономические подразделения могут выявляться при дополнительном введении в схему классификации других ландшафтных признаков и параметров, характеризующих геофильтрационные свойства вмещающих пород.

Как уже упоминалось, все естествоиспытатели стремятся к созданию "естественных" классификаций или хотя бы (в какой-то мере) к приближению этого вида классификаций. Пока еще никому не удалось дать логически удовлетворительное определение термина "естественной" классификации, как и выработать достаточно строгие ее критерии.

По мнению большинства специалистов, естественными считаются только те классификации, которые выражают законы природы, а некоторые исследователи (например, В.Ю. Забродин и А.А. Любищев /Забродин, 1981а/ допускают мысль, что формой закона может быть классификация, подтверждая это положение ссылками на Периодическую систему элементов Менделеева и кристаллографическую систематику Федорова – Шенфлиса.

В работе В.Ю. Забродина /1981а/ дается подробный анализ существующих критериев естественности, которые подразделяются на сильные и слабые.

Сильным критериям удовлетворяют только полностью естественные классификации, а слабым – классификации, занимающие промежуточное положение между полностью искусственными и естественными. Анализ приведенных критериев в цитируемой выше работе показывает, что они трудно поддаются как формализации, так и практическому применению. По-видимому, большинство этих критериев довольно сложно применить на практике вследствие того, что мы не обладаем достаточным количеством достоверной информации о природных объектах. В качестве примера можно привести второй критерий Уэлвелла, с помощью которого утверждается, что естественной будет такая классификация, когда упорядоченность в ней сохранится при смене классификационных признаков. Аналогичен этому критерию по своей сути критерий Мейена – Шрейдера, но последний дополняется такими требованиями, как объективность, надежность, прогностическая сила и др. Интересен, на наш взгляд, первый критерий Уэлвелла в интерпретации В.Ю. Забродина, утверждающий, что классификация тем естественней, чем более существенные связи она отражает. Данный критерий В.Ю. Забродин /1981а/ по своей шкале отнес наиболее близко к сильному критерию.

Предлагаемые в монографии схемы районирования бассейна оз. Байкал по условиям формирования естественных ресурсов подземных вод лишь отчасти удовлетворяют приведенным критериям, но основные функции классификаций – познавательную и прогностическую – они безусловно выполняют.

ЗАКЛЮЧЕНИЕ

Приведенные в монографии модели раскрывают далеко не все возможности эффективного использования аппарата многомерного анализа при решении гидрогеологических задач. Но такой цели и не ставилось, а стремление обосновать на конкретном природном объекте возможности применения этого арсенала статистических методов может быть вполне оправдано.

На базе построения многомерных моделей формирования ресурсов подземных вод горно-складчатых областей в монографии сделана попытка охватить два генеральных направления.

Первое – с позиции системного подхода разобраться в комплексе сложных структурных взаимосвязей элементов гидрогеологических систем с целью выделения групп информативных признаков. По существу, этот этап включает, кроме построения факторной модели, создание концептуальной основы взаимодействия элементов гидрогеологических систем.

Второе – построение многомерных классификаций гидрогеологических объектов по комплексу информативных признаков и использование процедур классификаций как методов познания процессов формирования ресурсов подземных вод.

При построении многомерных классификаций этих процессов весьма перспективен метод итерации, который заключается в последовательном разбиении исходной совокупности объектов на мелкие подмножества и предусматривает обратный путь – объединение этих объектов в множества при различных иерархических уровнях. Такие итерационные процедуры сопровождаются не только анализом динамики формирования объектов в классы, но и анализом соотношения признакового пространства классифицируемых объектов. Иллюстрация итерационных процедур классификаций достигается методами кластерного и факторного анализов.

Многовариантность классификаций, возникающая при использовании различных алгоритмов и переменной совокупности признаков и объектов, не может явиться серьезным препятствием для решения поставленных задач. Наоборот, в этом случае каждая классификация – средство к достижению поставленной цели, поскольку каждый вариант несет определенную информацию об интересующих нас процессах. Главный принцип, который здесь должен сохраняться, это комплексирование процедуры типологизации с другими методами статистической обработки данных, начиная с вычислений простых статистик для каждого класса и всей совокупности данных, вплоть до методов факторного анализа. Кроме сказанного, необходимым этапом представляется верификация таксономических алгоритмов на искусственных эталонных объектах. Проведение таких процедур позволяет из многообразия существующих алгоритмов автоматической классификации выбрать тот необходимый их набор, который даст максимальный эффект.

Указанные приемы позволят в значительной мере избежать формализма, который присущ большинству алгоритмов. В альтернативных условиях предпочтение отдается, как правило, той классификации, которая наиболее просто интерпретируется на основе принципов построения понятийных моделей.

Использование многомерных классификаций как одного из ведущих методов исследования пространственной неоднородности формирования ресурсов подземных вод в горно-складчатых областях может быть оправдано следующими основными теоретическими представлениями.

1. Многокомпонентностью структуры гидрогеологических систем, под структурой которых можно понимать множество форм координации составных частей (элементов) системы.

2. Стохастической (диффузной) природой процессов, протекающих в гидрогеологических системах.

3. Единством внутренних и внешних связей элементов системы и всей системы в целом.

В соотношении операций гидрогеологического районирования по принципу квазиоднородности гидрогеологических систем (любого порядка) и построении многомерных классификаций можно найти много общего. Конечная цель этих процедур – необходимая форма выражения наших знаний о формально-содержательной, логической, метрической и алгоритмической сторонах гидрогеологических систем.

Как показал опыт исследований, различия между понятийными и классификационными моделями в территориальном и количественном отношении (несоответствие числа районов и классов) чаще всего возникают вследствие использования неформальных, строго не разработанных принципов гидрогеологического районирования, в которых интуиция часто преобладает над фактами.

При соблюдении показанных в монографии логических процедур классификаций в рамках системного подхода можно добиться определенной адекватности схем районирования, рассматривая последнее как одну из задач классификации, поскольку практически гидрогеологическое районирование сводится к переработке информации о гидрогеологических системах.

Вполне естественно, что более строгие классификации (таксономические модели) базируются на использовании информации, основанной на метрических данных, а ее всесторонний анализ – залог успеха, который определяет практическую ценность многомерных классификаций и, следовательно, районирования.

Следует подчеркнуть, что проведение строгой классификации объектов должно сопровождаться классификацией их свойств (признаков). В монографии этому положению придается не менее важное значение, чем классификации самих объектов. Классификация признаков может рекомендоваться как перед началом классификации объектов, так и на всех ее этапах. Число признаков при этом может существенно меняться при движении от этапа к этапу с учетом поставленных задач.

Применяемый в работе комплекс многомерных методов базируется преимущественно на линейных связях между признаками. Может возникнуть вопрос – как поступить в случае нелинейных зависимостей? Область применения линейных моделей может быть существенно расширена за счет приведения к линейной форме таких нелинейных моделей, как мультипликативные, степенные, экспоненциальные и др. Способов минерализации нелинейных зависимостей разработано немало. Достоинство линейных моделей – относительная простота их интерпретации, а также доступность практического использования.

Хотелось бы отметить еще одно из основных достоинств многомерных классификаций при исследовании гидрогеологических процессов. Классификации являются не только результатом наших представлений о соотношении элементов гидрогеологических систем, но и дают начало новым направлениям по их исследованию.

Автор искренне признателен доктору геол.-мин. наук И.С. Ломоносову за научное редактирование монографии, докторам геол.-мин. наук А.И. Гаврищину, Н.М. Фролову, кандидату геол.-мин. наук С.М. Чесалову за ценные советы, полученные при работе над книгой, А.Н. Судакову и В.Н. Евдокимовой за помощь в трудоемких расчетах, а также благодарен Н.Г. Балуковой, Н.Г. Крашенинниковой и В.В. Романовой за помощь при оформлении рукописи к изданию.

ЛИТЕРАТУРА

Абрамович И.И., Дуденко Л.Н., Мишин В.И. О кластеризации магматических пород по петрохимическим данным. – В кн.: Математические методы в геологии. Л.: изд. ВСЕГЕИ, 1975, с. 83-92.

Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений.– М.: Статистика, 1974. – 240 с.

Андерсон Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963.– 500 с.

Аркадьев А.Г., Браверман Э.М. Обучение машины классификации объектов. – М.: Наука, 1971. – 192 с.

Арманд А.Д. Информационные модели природных комплексов. – М.: Наука, 1975. – 126 с.

Афанасьев А.Н. Водные ресурсы и водный баланс бассейна оз. Байкал. – Новосибирск: Наука. Сиб. отд-ние, 1976. – 237 с.

Белонин М.Д., Голубева В.А., Скублов Г.Т. Факторный анализ в геологии. – М.: Недра, 1982. – 269 с.

Беркин Н.С. Расчет некоторых элементов теплового и водного балансов для территории Прибайкалья (Иркутская область). – В кн.: Тепловой и водный режим некоторых районов Сибири. Л.: Наука. Ленингр. отд-ние, 1970, с. 43-57.

Беркин Н.С., Малий В.А. Опыт оценки балансовым методом естественных ресурсов артезианских вод Иркутского амфитеатра. – В кн.: Вопросы изучения и прогноза режима подземных вод. Вып. 25. М.: изд. ВСЕГИНГЕО, 1970, с. 103-110.

Боннер Р.Е. Некоторые методы классификации. – В кн.: Автоматический анализ сложных изображений. М.: Мир, 1969, с. 205-234.

Бугаец А.Н., Дуденко Л.Н. Математические методы при прогнозировании месторождений полезных ископаемых. – Л.: Недра, 1976. – 270 с.

Булавко А.Г. Водный баланс речных водосборов. – Л.: Гидрометеоиздат, 1971. – 303 с.

Ватанабе С., Ламберт П., Куликовский К. и др. Оценка и отбор параметров в задачах распознавания образов. – В кн.: Автоматический анализ сложных изображений. М.: Мир, 1969, с. 270-295.

Воронин Ю.А. Введение в теорию классификаций. – Новосибирск: изд. ВЦ СО АН СССР, 1982. – 194 с.

Всеволожский В.А. Ресурсы подземных вод южной части Западно-Сибирской низменности.– М.: Недра, 1973. – 88 с.

Всеволожский В.А. Подземный сток и водный баланс платформенных структур. – М.: Недра, 1983. – 167 с.

Гавич И.К. Теория и практика применения моделирования в гидрогеологии. – М.: Недра, 1980. – 355 с.

Гавич И.К. О некоторых понятиях и определениях в гидрогеологии. – Вод. ресурсы, 1984, № 5, с. 147-150.

Гавришин А.И. Оценка и контроль качества геохимической информации. – М.: Недра, 1980а. – 286 с.

Гавришин А.И. Использование искусственных эталонов для оценки методов классификации природных объектов. – В кн.: Оценка качества гидрогеологической и инженерно-геологической информации. Новочеркасск: изд. НПИ, 1980б, с. 100-109.

Гандин Л.С., Каган Р.С. Статистические методы интерпретации метеорологических данных. – Л.: Гидрометеоиздат, 1976. – 356 с.

Гапонцев Г.П. Оценка эффективности некоторых методов изучения однородности выборочных совокупностей на примере искусственных эталонов. – В кн.: Оценка качества гидрогеологической и инженерно-геологической информации. Новочеркасск: изд. НПИ, 1980, с. 109-115.

Гарцман И.Н., Карасев М.С., Лобанова Н.И. О густоте речной сети как индикаторе взаимосвязи грунтовых и поверхностных вод. – В кн.: Взаимосвязь поверхностных и подземных вод. М.: Изд-во МГУ, 1973, с. 112-114.

Гзовский М.В. Математика в геотектонике. – М.: Недра, 1971. – 240 с.

Гзовский М.В. *Основы тектонофизики.* – М.: *Наука*, 1975. – 553 с.

Гидрологическая изученность. Забайкалье. Т. 16, вып. 3. – Л.: Гидрометеоиздат, 1966.– 159 с.

Григорьева Н.В., Чесалов С.М. Гидрогеохимические исследования при обосновании разведки эксплуатационных запасов сульфидных вод Кемери. – В кн.: Вопросы гидрогеологии. Закономерности формирования и оценка ресурсов подземных вод. М.: Изд-во МГУ, 1981, с. 183-194.

Двинских Н.И. Программа статистического анализа полей и процессов. – В кн.: Исследования по геомагнетизму, аэрономии и физике Солнца. Вып. 33. Иркутск: Тр. СибИЗМИР СО АН СССР, 1975, с. 137-143.

Девис Дж. Статистика и анализ геологических данных. – М.: Мир, 1977. – 572 с.

Джефферс Дж. Введение в системный анализ: применение в экологии. – М.: Мир, 1981.– 256 с.

Дружинин В.В., Конторов Д.С. Проблемы системологии. – М.: Сов. радио, 1976. – 296 с.

Дубинчук В.Т. Гидрогеологическое прогнозирование с позиций системного анализа. – В кн.: Формирование подземных вод как основа гидрогеологических прогнозов. Т. 1. М.: Наука, 1982, с. 92-97.

Дюран Б., Оделл П. Кластерный анализ. – М.: Статистика, 1977. – 127 с.

Евдокимова В.Н. Автоматизированная система обработки геолого-геохимической информации методом многомерных полей. – В кн.: Геохимические методы поиска рудных месторождений в Сибири и на Дальнем Востоке. Новосибирск: Наука. Сиб. отд-ние, с. 3-26.

Елисеева И.Н., Руковишников В.О. Группировка, корреляция, распознавание образов. – М.: *Статистика*, 1977. – 144 с.

Естественные ресурсы подземных вод юга Восточной Сибири. – Новосибирск: Наука. Сиб. отд-ние, 1976. – 127 с.

Жуков Р.А. Системный подход и методологические резервы теоретической геологии. – В кн.: *Методы теоретической геологии. Л.: Недра*, 1978, с. 24-80.

Забродин В.Ю. О критериях естественности классификаций. – Бюл. НТИ. Сер. 2. Геол. и полезные ископаемые, 1981а, № 8, с. 22-24.

Забродин В.Ю. Системный анализ дизъюнктов. – М.: Наука, 1981б. – 200 с.

Зальцберг Э.А. Статистические методы прогноза естественного режима уровня грунтовых вод. – Л.: Недра, 1976. – 101 с.

Зекцер И.С. Закономерности формирования подземного стока и научно-методические основы его изучения. – М.: Наука, 1977. – 172 с.

Иберла К. Факторный анализ. – М.: Статистика, 1980. – 298 с.

Йёреског К.Г., Клован Д.И., Реймент Д.И. Геологический факторный анализ. – Л.: Недра, 1980. – 223 с.

Караванов К.П. Бассейны подземных вод горно-складчатых областей Восточной Азии.– М.: Наука, 1977. – 144 с.

Караванов К.П. Типизация подземных водоносных систем. – М.: Наука, 1980. – 152 с.

Ковалевский В.С. Многолетние колебания уровней подземных вод и подземного стока.– М.: Наука, 1976. – 270 с.

Ковалевский В.С. Многолетняя изменчивость ресурсов подземных вод. – М.: Наука, 1983. – 205 с.

Кожара В.Л. Как объективизировать процедуру естественной классификации природных [illegible]д. – В кн.: *Проблемы теоретической и региональной гидрогеохимии. М.: Изд-во* МГУ, 1979, с. 97-100.

Кожара В.Л. Функции классификаций. – В кн.: Теория классификаций и анализ данных. Новосибирск: изд. ВЦ СО АН СССР, 1982, с. 5-19.

Кондаков Н.И. Логический словарь. – М.: Наука, 1971, с. 214-216.

Коноплянцев А.А., Семенов С.М. Прогноз и картирование режима грунтовых вод. – М.: Недра, 1974. – 214 с.

Коротков А.И. Гидрогеохимический анализ при региональных геологических и гидрогеохимических исследованиях. – Л.: Недра, 1983. – 231 с.
Косыгин Ю.А., Соловьев В.А. Статические, динамические и ретроспективные системы в геологических исследованиях. – Изв. АН СССР. Сер. геол., 1969, № 6, с. 9-17.
Крамбейн У., Грейбилл Ф. Статистические модели в геологии. – М.: Мир, 1969. – 397 с
Крамбейн У., Кауфман М., Мак-Кеммон Р. Модели геологических процессов. – М.: Мир, 1973. – 150 с.
Кроули У. FLAG-свободно-лагранжев метод для численного моделирования гидродинамических течений в двух измерениях. – В кн.: Численные методы в механике жидкости. М.: Мир, 1973, с. 135-146.
Куделин Б.И. Принципы региональной оценки естественных ресурсов подземных вод. – М.: Изд-во МГУ, 1960. – 344 с.
Куделин Б.И., Фиделли И.Ф. К вопросу о гидрогеологическом районировании территории СССР. – В кн.: Пятая научная отчетная конференция геологического факультета МГУ 16-19 марта 1970 г. (Тез. докл.). М.: Изд-во МГУ, 1970, с. 270-272.
Леви К.Г. Вертикальные движения земной коры в Байкальской рифтовой зоне. – В кн.: Проблемы разломной тектоники. Новосибирск: Наука. Сиб. отд-ние, 1981, с. 142-170.
Лаудон Т. ЭВМ и машинные методы в геологии. – М.: Мир, 1981. – 318 с.
Лисенков А.Б., Денисов В.А. Генетические связи элементов химического состава подземных вод южного Мангышлака, изученные с помощью факторного анализа. – Изв. вузов. Геол. и разв., 1977, № 8, с. 83-92.
Лоули Д., Максвелл А. Факторный анализ как статистический метод. – М.: Мир, 1967.– 141 с.
Мазуров В.Д., Казанцев В.Ф., Белецкий Н.Г. и др. Пакет КВАЗАР прикладных программ распознавания образов. Версия 2. Информационные материалы по матобеспечению. – Свердловск: изд. УНЦ АН СССР, 1970. – 119 с.
Марчук Г.И., Агошков В.И. Введение в проекционно-сеточные методы. – М.: Наука, 1981. – 416 с.
Мезенцев В.С. Расчеты водного баланса. – Омск: изд. Омск. СХИ, 1973. – 76 с.
Методы теоретической геологии. – Л.: Недра, 1978. – 336 с.
Миллер Р., Кан Дж. Статистический анализ в геологических науках. – М.: Мир, 1965.
Налимов В.В. Теория эксперимента. – М.: Наука, 1971. – 207 с.
Николаев Ю.В. Классификация гидрометеорологических процессов с помощью ЭВМ. – Л.: Гидрометеоиздат, 1976. – 34 с.
Огильви Н.А. Физические и геологические поля в гидрогеологии. – М.: Наука, 1974.– 160 с.
Окунь Я. Факторный анализ. – М.: Статистика, 1974. – 200 с.
Оноприенко В.И. Взаимосвязь структурного, генетического и системного подхода в геологических исследованиях. – В кн.: Методологические вопросы геологических наук. Киев: Наукова думка, 1974, с. 67-81.
Основы гидрогеологии. Общая гидрогеология. – Новосибирск: Наука. Сиб. отд-ние, 1980. – 231 с.
Павличенко Л.М., Шапиро С.М. Исследования процессов формирования гидрогеохимических условий древней дельты р. Или методом главных компонент. – Вод. ресурсы, 1982, № 4, с. 77-84.
Павлов А.Н. Основы системного подхода в геологии. – Л.: изд. ЛГМИ, 1981. – 84 с.
Пиннекер Е.В., Писарский Б.И., Ломоносов И.С. и др. Гидрогеология Прибайкалья. – М.: Наука, 1968. – 170 с.
Пиннекер Е.В. Проблемы региональной гидрогеологии. – М.: Наука, 1977. – 196 с.
Писарский Б.И. Особенности формирования подземных вод Байкальского артезианского бассейна. – В кн.: Подземные воды Сибири и Дальнего Востока. М.: Наука, 1971, с. 59-69.
Писарский Б.И. Подземный сток бассейна оз. Байкал. – В кн.: Очерки по гидрогеологии Сибири. Новосибирск: Наука. Сиб. отд-ние, 1973, с. 43-52.
Писарский Б.И. Формирование подземного стока в тектонически активных областях. – В кн.: Формирование подземных вод как основа гидрогеологических прогнозов. Т. 1 (Матер. первой Всесоюз. гидрогеол. конф.). М.: Наука, 1982, с. 248-250.
Писарский Б.И., Русанов Е.В. Формирование подземного стока в бассейнах Голоустной и Бугульдейки. – В кн.: Круговорот вещества и энергии в водоемах. Вып. IV (Тез. докл. Всесоюз. лимнол. совещ.). Иркутск, 1981, с. 57-58.

Писарский Б.И., Хаустов А.П. Применение гидрохимических методов для оценки взаимодействия поверхностных и подземных вод в горно-складчатых областях на примере бассейна оз. Байкал. – В кн.: Взаимосвязь поверхностных и подземных вод. М.: Изд-во МГУ, 1973, с. 211-213.

Писарский Б.И., Хаустов А.П. Поясность подземного стока в горно-складчатых областях основа региональной оценки и картирования естественных ресурсов подземных вод.– В кн.: Методы оценки ресурсов подземных вод. М.: Наука, 1979, с. 78-80.

Писарский Б.И., Хаустов А.П. Применение факторного анализа для районирования горно-складчатых областей по условиям формирования подземного стока. – Вод. ресурсы, 1982, № 4, с. 69-76.

Писарский Б.И., Хаустов А.П., Костин В.И. Поясность подземного стока в горно-складчатых областях на примере бассейна р. Джида в Забайкалье. – Вод. ресурсы, 1975, № 5, с. 100-106.

Плюта В. Сравнительный многомерный анализ в экономических исследованиях. – М.: Статистика, 1980. – 114 с.

Подземный сток на территории СССР. – М.: Изд-во МГУ, 1966. – 302 с.

Попов О.В. Подземное питание рек. – Л.: Гидрометеоиздат, 1968. – 292 с.

Ресурсы поверхностных вод СССР. Т. 16, вып. 2. – Л.: Гидрометеоиздат, 1972. – 595 с.

Ресурсы поверхностных вод СССР. Основные гидрологические характеристики. Ангаро-Енисейский район. Забайкалье. Т. 16, вып. 3. – Л.: Гидрометеоиздат, 1976. – 178 с.

Ржаницын Н.А. Морфологические и гидрологические закономерности строения речной сети. – Л.: Гидрометеоиздат, 1960. – 238 с.

Роговская Н.В. Гидрогеологическое картирование (обзор отечественного и зарубежного опыта составления гидрогеологических карт). – М.: Наука, 1981. – 132 с.

Роговская Н.В., Кунин В.Н. О международной гидрогеологической карте. – Сов. геология, 1969, № 11, с. 140-143.

Садовский В.Н. Основания общей теории систем. – М.: Наука, 1974. – 279 с.

Сборник научных программ на ФОРТРАНе. Вып. 2. – М.: Статистика, 1974. – 160 с.

Сиськов В.И. Корреляционный анализ в экономических исследованиях. – М.: Статистика, 1975. – 117 с.

Сокал Р.Р. *Кластер-анализ и классификация: предпосылки и основные направления.* – В кн.: Классификация и кластер. – М.: Мир, 1980, с. 7-19.

Соколов Б.Л., Саркисян В.О. Подземное питание горных рек. – Л.: Гидрометеоиздат, 1980. – 240 с.

Смирнов Б.И. Корреляционные методы при парагенетическом анализе. – М.: Недра, 1981. – 176 с.

Справочник по климату СССР. Иркутская область. Вып. 22, ч. IV. – Л.: Гидрометеоиздат, 1968. – 160 с.

Справочник по климату СССР. Бурятская АССР. Читинская область. Вып. 23, ч. IV. – Л.: Гидрометеоиздат, 1968. – 185 с.

Степанов В.М. Гидрогеологические структуры Забайкалья. – М.: Недра, 1980. – 176 с.

Судаков А.Н., Хаустов А.П., Писарский Б.И. Моделирование теплового поля месторождения Шаргалжут (гидротермы Хангая в МНР). – В кн.: Применение математических методов и ЭВМ в геологии. (Тез. докл. к третьему областному семинару в г. Новочеркасске 24-26 июня 1983 г.). Новочеркасск, 1983, с. 87-88.

Типология и классификация в социологических исследованиях. – М.: Наука, 1982. – 295 с.

Уемов А.И. Системный подход и общая теория систем. – М.: Мысль, 1978. – 272 с.

Умаров У. Способы представления гидрогеологических объектов и их связей как систем.– В кн.: Формирование подземных вод как основа гидрогеологических прогнозов. Т. 1. (Матер. первой Всесоюз. гидрогеол. конф.). М.: Наука, 1982, с. 165-169.

Федоров В.Н., Хаустов А.П. К расчету подземной составляющей речного стока по гидрохимическим данным. – В кн.: Круговорот вещества и энергии в озерах и водохранилищах. Листвиничное на Байкале. Т. 1, 1973, с. 184-186.

Фиделли И.Ф. Особенности формирования и распределения подземного стока на территории горно-складчатых областей. – В кн.: Научные основы изучения и охраны подземных вод. Ч. 2. М.: Изд-во МГУ, 1980, с. 72-80.

Флоренсов Н.А. Мезозойские и кайнозойские впадины Прибайкалья. – М.-Л.: Изд-во АН СССР, 1960. – 257 с.

Фролов Н.М. Основы гидрогеологической стратификации. – Изв. АН СССР. Сер. геол., 1981, № 11, с. 9-20.

Фролов Н.М. Упорядочение понятийно-терминологической базы в гидрогеологии. – Вод. ресурсы, 1984, № 3, с. 73-90.

Харман Г. Современный факторный анализ. – М.: Статистика, 1972. – 485 с.

Хаустов А.П. Использование факторной модели для анализа фильтрационных свойств водовмещающих пород (на примере бассейна р. Джиды). – В кн.: Региональная гидрогеология и инженерная геология Восточной Сибири. Новосибирск: Наука. Сиб. отд-ние, 1978, с. 13-16.

Хаустов А.П. Временная изменчивость подземного стока Юго-Западного Забайкалья. – В кн.: Подземный сток на территории Сибири и методы его изучения. Новосибирск: Наука. Сиб. отд-ние, 1979а, с. 92-96.

Хаустов А.П. О применении некоторых методов многомерного анализа в гидрогеологии. – В кн.: Тезисы докладов девятого совещания по подземным водам Сибири и Дальнего Востока (Всесоюз. совещ. по подзем. водам Востока СССР). Иркутск – Петропавловск-Камчатский, 1979б, с. 17-18.

Хаустов А.П. Роль атмосферных осадков в формировании подземного стока Юго-Западного Забайкалья. – В кн.: Вопросы изучения поверхностных и подземных вод Сибири. Иркутск: Изд-во Иркут. ун-та, 1981, с. 86-90.

Хаустов А.П. Закономерности формирования подземного стока и методы его оценки. – Иркутск: Изд-во Иркут. ин-та, 1982а. – 80 с.

Хаустов А.П. Системный подход к анализу закономерностей формирования ресурсов подземных вод в горно-складчатых областях. – В кн.: Формирование подземных вод как основа гидрогеологических прогнозов. Т. 2. М.: Наука, 1982б, с. 72-75.

Хаустов А.П. Гидрохимический баланс рек бассейна оз. Байкал. – В кн.: Комплексные исследования водных ресурсов Сибири. Вып. 56. (Тр. ЗапСибНИИ Госкомгидромета). М.: Гидрометеоиздат, 1982в, с. 87-97.

Хаустов А.П. Факторная модель формирования подземного стока в бассейне оз. Байкал. – В кн.: Гидрология Байкала и других водоемов. Новосибирск: Наука. Сиб. отд-ние, 1984а, с. 129-138.

Хаустов А.П. Классификация условий формирования естественных ресурсов подземных вод в бассейне оз. Байкал методом многомерных полей. – В кн.: Гидрогеологические исследования восточных районов СССР и некоторых стран Азии. Иркутск, 1984б, с. 98-111.

Хаустов А.П., Рассохин В.М. *Роль атмосферных осадков в формировании ионного стока* рек бассейна оз. Байкал. – В кн.: Гидрохимические материалы. Т. 80. Л.: Гидрометеоиздат, 1982, с. 47-54.

Хаустов А.П., Федоров В.Н. *К анализу полей гидрогеологических характеристик реки* Джиды. – В кн.: Гидрогеологические исследования в Приамурье. Владивосток: изд. ДВНЦ АН СССР, 1979, с. 20-23.

Хаустов А.П., Чесалов С.М. Районирование бассейна оз. Байкал по условиям формирования подземного стока методом факторного анализа. – В кн.: Круговорот вещества и энергии в водоемах. Вып. IV. (Тез. докл. на Всесоюз. лимнол. совещ.). Иркутск, 1981, с. 54-55.

Хортон Р.Е. Эрозионное развитие рек и водосборных бассейнов. – М.: Изд-во иностр. лит., 1948. – 158 с.

Чесалов С.М. Классификация искусственных эталонов объектов с помощью факторного анализа. – В кн.: Оценка качества гидрогеологической и инженерно-геологической информации. Новочеркасск: изд. НПИ, 1980, с. 120-124.

Шапиро С.М., Павличенко Л.М., Подольный О.В. Гидрогеологические прогнозы подземного стока в озеро Балхаш. – Алма-Ата: Наука, 1982. – 128 с.

Швецов П.Ф. К системе гидрогеологических понятий. – Вод. ресурсы, 1984, № 3, с. 91-95.

Швецов П.Ф., Киселев П.А. К систематизации объектов региональной гидрогеологии. – Бюл. МОИП. Отд. геол., 1979, т. 54, вып. 1, с. 96-110.

Шенькман Б.М. Картирование подземного стока на основе зимних гидрометрических исследований. – В кн.: Региональная гидрогеология и инженерная геология Восточной Сибири. Новосибирск: Наука. Сиб. отд-ние, 1978, с. 136-150.

Шерман С.И. Физические закономерности развития разломов земной коры. – Новосибирск: Наука. Сиб. отд-ние, 1977. – 127 с.

Шмагин Б.А. Вопросы исследования распределения величин подземного стока вероятностно-статистическими моделями и оценками. – В кн.: Вопросы гидрогеологии. Закономерности формирования и оценка ресурсов подземных вод. М.: Изд-во МГУ, 1981, с. 54-62.

Юдин Б.Г. Становление и характер системной ориентации. – В кн.: Системные исследования. Ежегодник 1971. М.: Наука, 1972, с. 18-35.

Ясько В.Г. Подземные воды межгорных впадин Забайкалья. – Новосибирск: Наука. Сиб. отд-ние, 1982. – 169 с.

Angewandte Hydrologie. Teil 1 Berechnung und Regelung des Dürchflusses der Flüsse. – Berlin: Verlag für Bauwesen, 1980. – 528 S.

Angewandte Hydrologie. Der Wasserhaushalt der Flüssgebiete. – Berlin: Verlag für Bauwesen, 1980, Teil 2. – 542 S.

Gabriel K.B. Simultaneous test procedures in multivariate analysis of ,Variance. – J. Biometrika, 55, 1966, p. 489–504.

Jahn W., Vahle H. Die Faktoranlyse und ihre Anwendunge. – Berlin: Verlag die Wirtschaft, 1970. – 228 S.

Luckner L. Einige Aspekt für den Einsatz einer rationalen Simulatinstechnik zur Steuerung Unterirdischer Strömungsvorgänge. – Wissenschaftliche Zeitschrift der Technischen Universität, Dresden, 1975, Bd 24, H. 6, S. 1437-1439.

Pisarskii B.I., Khaustov A.P. The Role of Precipitation in the Formation of Saline Groundwater Runoff from Rivers in the Baikal Area. – J. Water guality Bulletin, 1981, v. 6, 1, p. 24-25.

Sneath P., Sokal R.R. Principles of Numerical Taxonomy. – N.Y.: Freeman and Cooper, 1973. – 553 p.

Temple I.T. The Use of Facktor Analysis in Geology. – I. of Iust. Assoc. of Math. Geology, 1978, v. 10, 4, p. 379-378.

Weber E. Einführung in die Facktorenanaliyse. – Iena: Verlag, Gustav Fischer, 1974. – 192 S.

Harfe I., Sponholz G. Komplexe Bearbeitung ingenieurgeologischer Daten mit Klassifizierungs – und Erkennungsverfahren. – Zeitschrift für angewandte Geologie, 1975, Bd 21, H. 8, S. 372-376.

ОГЛАВЛЕНИЕ

От редактора 3

Глава 1. Системный подход и методологические аспекты классификации гидрогеологических систем 5

1.1. *Системно-кибернетический принцип представления гидрогеологических систем* . . . . . –

1.2. Методологические аспекты типологии и классификации гидрогеологических систем 19

Глава 2. Методика расчета признакового пространства многомерных моделей формирования ресурсов подземных вод 26

2.1. Основные параметры подземного стока 28

2.2. Геолого-структурные признаки 29

2.3. Морфометрические и геоморфологические признаки 30

2.4. Гидрометеорологические признаки 31

2.5. Признаки высотной поясности подземного стока 35

Глава 3. Факторные модели формирования естественных ресурсов подземных вод горно-складчатых областей 36

3.1. Методология применения и основы факторного анализа . . . . . –

3.2. R-факторная модель формирования ресурсов подземных вод 44

3.3. Q-факторная модель формирования ресурсов подземных вод 71

Глава 4. Автоматическая классификация условий формирования ресурсов подземных вод 76

4.1. Общие положения задач автоматической классификации . . . . . –

4.2. Классификация условий формирования ресурсов подземных вод на основе иерархической таксономии 81

4.3. Классификация условий формирования ресурсов подземных вод на основе неиерархической таксономии 88

4.4. Оценка эффективности и сравнительный анализ методов классификации 95

Заключение 103

Литература 105

Александр Петрович Хаустов

МНОГОМЕРНЫЙ АНАЛИЗ
ГИДРОГЕОЛОГИЧЕСКИХ СИСТЕМ
ГОРНО-СКЛАДЧАТЫХ ОБЛАСТЕЙ

Утверждено к печати
Институтом земной коры СО АН СССР

Редактор издательства Л.М. Акентьева

Художественный редактор М.Ф. Глазырина

Художник Н.А. Пискун

Технический редактор Н.М. Остроумова

Корректоры Е.В. Золина, Н.В. Лисина

ИБ № 30080

Сдано в набор 30.04.86. Подписано к печати 09.09.86.
МН -01253. Формат 70x100 1/16 Бумага офсетная.
Офсетная печать. Усл. печ. л. 9,1. Усл. кр.-отт. 9,4.
Уч.-изд.л. 9,5. Тираж 600 экз. Заказ № 301.
Цена 1 р. 40 к.

Ордена Трудового Красного Знамени издательство "Наука",
Сибирское отделение. 630099, Новосибирск, 99, Советская, 18.

4-я типография издательства "Наука". 630077, Новосибирск, 77,
Станиславского, 25.

1 р. 40 к.

4706

ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ